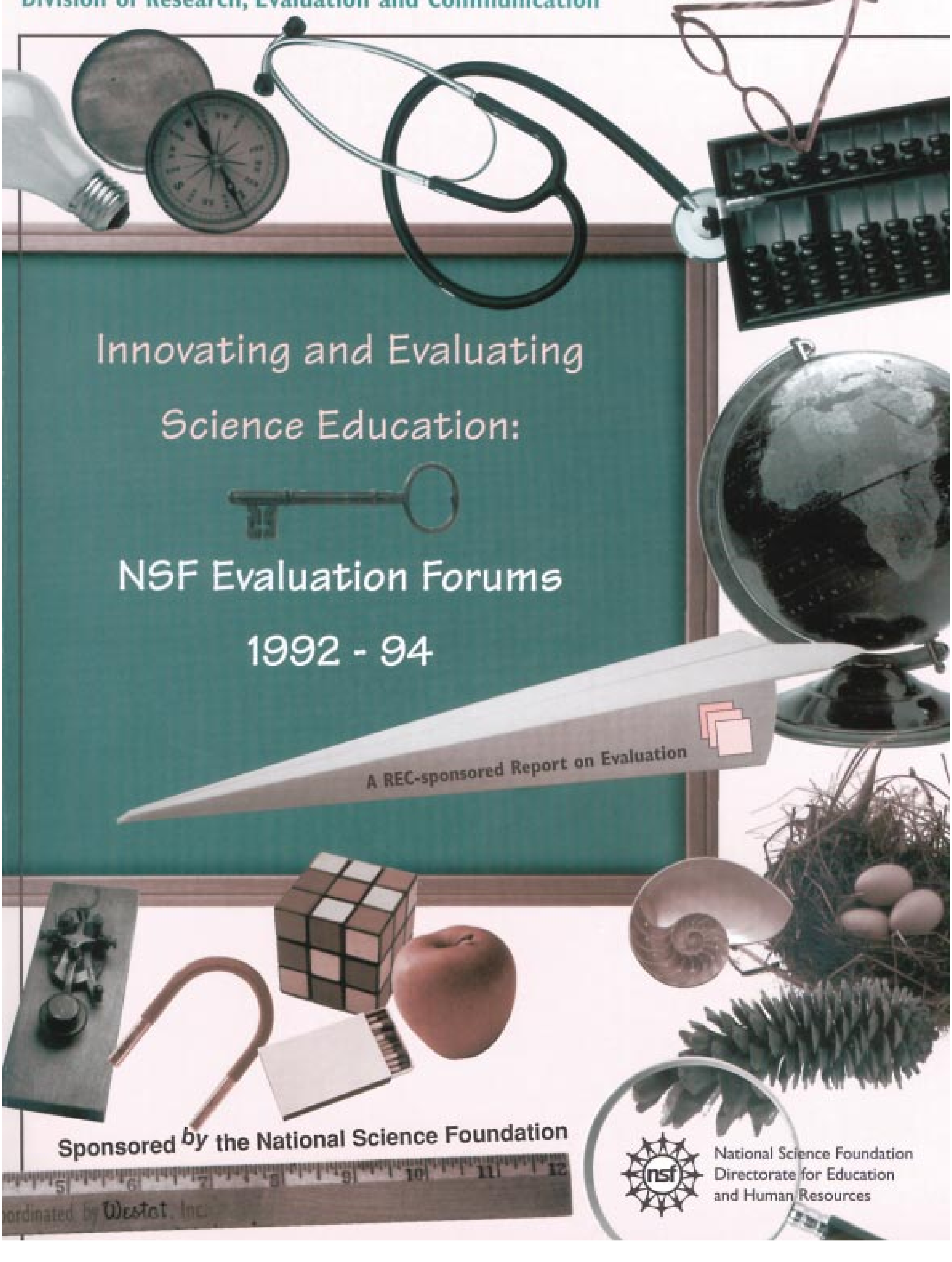
Division of Research, Evaluation and Communication



Innovating and Evaluating Science Education:

NSF Evaluation Forums 1992 - 94

A REC-sponsored Report on Evaluation

Sponsored by the National Science Foundation

National Science Foundation
Directorate for Education
and Human Resources

Coordinated by Westat, Inc.

INNOVATING AND EVALUATING SCIENCE EDUCATION: NSF EVALUATION FORUMS, 1992-94

Prepared under Contract
#SED 92-55369

by

Westat, Inc.

October 1995

NSF Program Officer
Floraline Stevens

Directorate for Education
and Human Resources

Division of Research,
Evaluation and Dissemination

Acknowledgments

Kenneth Travers, former Director of the National Science Foundation's Division of Research, Evaluation and Dissemination (RED), was the first to suggest that evaluation forums could be of benefit to those responsible for the evaluations of federally funded educational projects, especially NSF Program Directors in the Directorate for Education and Human Resources (EHR). His contribution to the formation and early implementation of the forums will long be remembered.

The NSF Program Directors who responded to the call for assistance in planning the forums developed the format for the program and suggested topics and speakers. Their resourcefulness and commitment contributed greatly to the success of the EHR Evaluation Forums. Many thanks to evaluation theorist Michael Scriven, who was the first speaker at the first EHR Forum.

Special thanks to Daryl Chubin, the current Director of RED, who saw the important contributions of the forums and gave his wholehearted support to continuing them.

Luther Williams, EHR Assistant Director, and his staff provided support and encouragement for the concept and implementation of the EHR Evaluation Forums, and the EHR Division Directors and their Program Directors showed their support by attending the forums for 2 years. Special mention must also be given to the loyal RED Program Directors who attended the forums and to those among them who volunteered to be speakers. It was their core support that made the forums successful.

Lastly, thanks to Joy Frechtling and Joan Michie of Westat, Inc., who handled the logistics of getting out-of-town speakers to and from the forums in good and bad weather. They made the task much easier for NSF staff.

— Floraline Stevens

Contents

Education Program Evaluation at NSF: What Difference Does It Make?

Daryl E. Chubin, Director
Division of Research, Evaluation and Dissemination
Directorate for Education and Human Resources
National Science Foundation

Introduction

Having completed a year at the National Science Foundation (NSF), I can now make some observations as an insider that were formerly those only of an outsider. This is a crucial distinction. In contrast, the authors of the papers that follow, originally presenters in a series of Evaluation Forums sponsored by NSF's Education and Human Resources Directorate (EHR), remain outside. That is their value, much like the value of doing program evaluation the way EHR does. It emphasizes independence, a focus on program objectives, and ultimately measurement of what difference the program has made amidst the panoply of efforts to improve the teaching and learning of mathematics and science across the education universe, that is, prekindergarten through grade 12 into undergraduate study, and graduate school, and early careers (see figure 1).

My remarks here serve two purposes: they sketch how EHR goes about its business of evaluation (a topic on which I have lectured extensively this past year),[1] and they illustrate the burden that grows on all who are charged with the responsibility of program evaluation, be they project directors, school administrators, analytical organizations working under contract to Federal agencies, or academic researchers conducting evaluations as expert consultants.

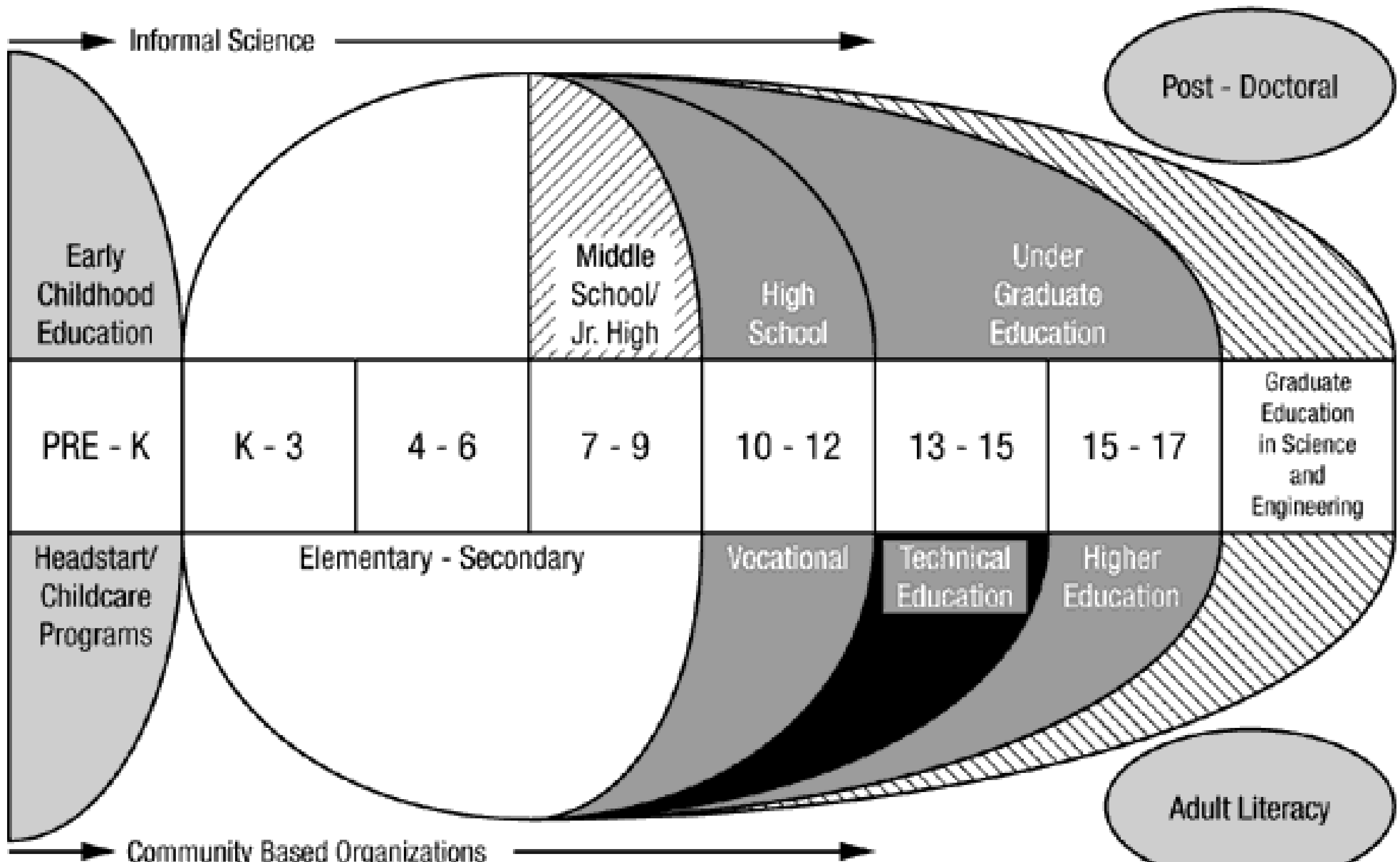
The "E" Word

Evaluation is no longer a dreaded word at EHR, but its mention still quickens the pulse of most program officers and those who must account for, budget, plan, and allocate scarce resources. The reason is that the Federal context for decision making, squeezed by reduced discretionary funding, has become oriented to outcomes. The National Performance Review, Government Performance and Results Act of 1993, and agency budget requests that must contain strategic plans all reflect a concern for performance, coordination, priority setting, and measures of returns on investment. These are all variations on the question, what difference did the program make? The shift from a preoccupation with inputs, i.e., how did you spend the program's monies, to outcomes is clear. It is also correct, in my view, and indeed overdue. And those who deem it a political gimmick that shall pass are fooling only themselves.

Program evaluation is on the cusp of these changes—cultural changes in the way Federal agencies intervene to enhance, expedite, improve, empower, and learn about what works in what settings. It is a tool for learning things that arguably are unknowable by any other means. So I take program evaluation as a unique input to, not a revelation about, decision making. If done in a competent and timely fashion, it can answer certain questions. It can also fail to answer others, which are more judgmental, such as, compared to what? and was it cost-effective? Such judgments belong to management with an intimate knowledge of the whole portfolio of activities in which a program is but one component.

Figure 1.

THE EDUCATION "UNIVERSE"



EHR's Model of Program Evaluation

We can ask three questions about EHR's program evaluation activity:

- Why do we do it?
- What do we do?
- How do we use the results?

These questions are summarized in the model depicted in figure 2.

Why evaluate? Simply put, NSF is *expected* to account to Congress and the Administration for its educational activities. What people tend to forget is that with EHR's rising appropriations, the number of strings attached to programs also grows. We get hundreds of inquiries annually demanding information on why a 2-year-old NSF program has not increased student test scores dramatically. Such queries attest to the short-term mentality that abounds and what accountability means in a Federal agency context.

But program evaluation is not done just for accountability purposes; it is also intended as a management and planning tool. To that end, we have developed (under contract) an EHR Impact Database that can be used by program officers for monitoring program inputs and process (as is currently done for the Experimental Program to Stimulate Competitive Research (EPSCoR) and the Alliances for Minority Participation program, as well as in evaluations of those programs.

What we do is put *all* EHR programs (currently there are 33) on a 5-year evaluation cycle, so at least one-third of the portfolio is being evaluated at any time. Most of these evaluations overlap 2 fiscal years and cost $100,000 to over $1 million to complete. (Definitions and the fiscal year 1994 schedule of EHR's evaluation activities, as well as a chart showing the contractors involved in evaluation efforts for the Division of Research, Evaluation and Dissemination, are appended to this paper.)

Some of these evaluations are shorter term and narrower in scope than full evaluations; we call them impact studies, and currently, three are under way: Young Scholars, Presidential Awards for Excellence in Science and Mathematics Teaching, and Alliances for Minority Participation.

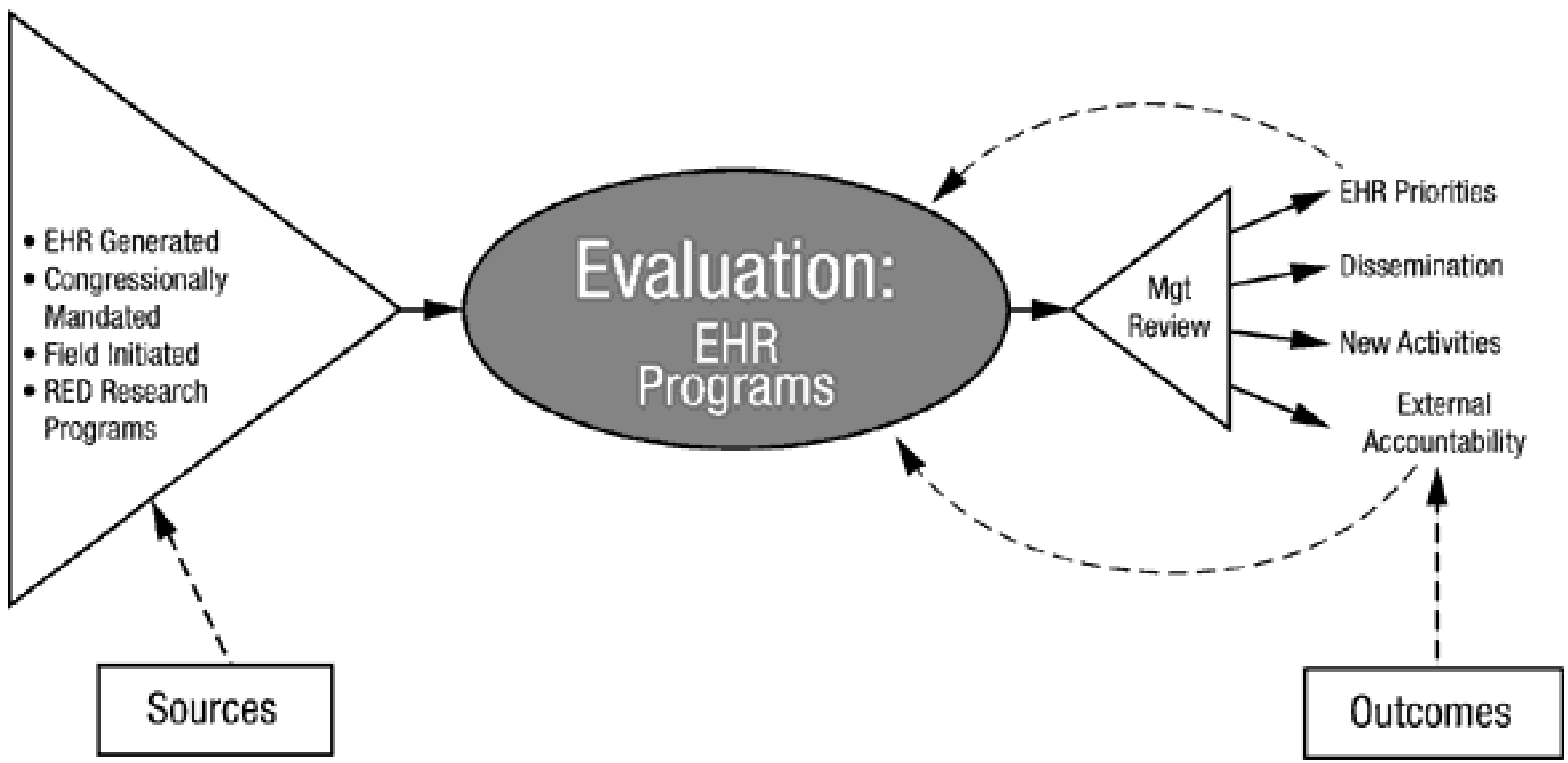
A staff of evaluation officers (four full-time-equivalent members) oversees the design and execution of evaluations by independent or third-party performers who compete for these contracts. In fiscal year 1994, program evaluation was about a $12 million EHR investment (representing roughly 2 percent of the Directorate's budget). This does not include project evaluations conducted by principal investigators or their designees in the field, a practice that has long been a part of projects supported by the Division of Elementary, Secondary, and Informal Education and the Division of Undergraduate Education.[2]

Beginning in fiscal year 1995, EHR will have five prime contractors and 15 subcontractors assisting in its program evaluation work (see the appendix). This is a daunting administrative task that interposes EHR evaluation officers as liaisons between the Program Officer whose program is being evaluated and the independent contractor. I consider this a fragile relationship. It entails keeping the principals in close touch, but at arm's length, so that the program's goals are captured without compromising the integrity of the evaluation—a constant challenge compounded by the extensive paperwork required by procurement procedures.

How do we plan to use the results of program evaluations? Very few evaluations have been completed to date. Twenty are in progress, so results are forthcoming. The evaluation staff and I will submit to the Office of the Assistant Director for Education and Human Resources the final contractor report plus a cover memorandum. This memo will suggest actions based on our interpretation of the findings on program "success" or "effectiveness" if the evaluation was of the summative variety; attention to "process" and issues encompassed in a formative evaluation will be similarly annotated.

Figure 2.

Model of EHR Program Evaluation



The point is that nothing automatically happens as a result of an EHR program evaluation. This is represented in figure 2 by the triangle marked management review. It is part of the process to decide what changes, if any, should be made to the program. This is the bottom line of the evaluation (see figure 3); it answers the big questions that management must answer—and act on. It helps the Assistant Director determine gaps in the EHR portfolio, identify candidates for consolidation, detect emerging priority areas that may warrant programmatic treatment, and, perhaps above all, indicates what should be disseminated to various offices as examples of best practice. Thus, dissemination is strongly coupled to evaluation as a mechanism of back-end quality control. It allows us to share quality processes and products for local adoption/ adaptation and may specify the need for the provision of technical assistance.[3]

Program Evaluation:

The Bottom Line

- Would any of the observed outcomes have occurred if this program did not exist?
- What is the "value added" by this program to the state of SMET education in the Nation?
- What lessons have been learned from the implementation of this program that could be applied to the development and implementation of other programs?

Figure 3.

Measuring Impacts

EHR staff (and particularly the Assistant Director) are constantly asked, "What difference does this program make?" There are as many answers as there are measures or indicators of progress, preparation, and achievement. Sometimes these measures of impacts are straightforward and easy to quantify, e.g., the number of minority students who complete a baccalaureate in science disciplines as a result of participating in an NSF-sponsored support program. If the slope is in the right direction, we think we are making the difference. (Other program data will help us fill in the inferential blanks.)

But measuring systemic change is particularly daunting. NSF has pioneered systemic education reform through the Statewide Systemic Initiatives (SSI), the Urban Systemic Initiatives, and now the Rural Systemic Initiatives and Local Systemic Initiatives programs.[4] An important, of-untold value of the SSI program is that it has afforded EHR the opportunity to ***invent systemic evaluation***. This means measuring not only what is occuring inside schools, but also their connection to other apparatus in the local and more distant environment, namely, negotiations with the political system to effect the conditions for change.

This is largely uncharted territory that goes beyond the familiar cry about raising student test scores. The year one evaluation report on SSI was published in June 1994.[5] It shows, to cite only a handful of findings, that

- Other sources are providing, in the aggregate, a **dollar-for-dollar match** to NSF's investment;
- Reform strategies vary greatly by state, but a shared emphasis has been on **inservice training** for teachers;

- The **absence of curriculum frame-works** (no doubt exacerbated by the lack of content standards in science) is impeding reform in some states; and

- **Public awareness** of K-12 reforms implemented in their states continues to lag (which stalls the momentum for change).

To me, this interim evaluation has indicated to NSF that the infrastructure has been created for mathematics and science reform, spurred in many states by the SSI. I believe this is a major impact of the program. In other states where reform has many sponsors, NSF has been an experimental force for the implementation and delivery of innovative mathematics and science at the elementary and secondary levels.

In sum, systemic reform seeks enduring impact in the teaching and learning of science and math. But the system has many working parts, and you cannot fix one without addressing the others (see figure 4).[6] Systemic evaluation must measure all parts. It must seek to capture how effective NSF has been as a change agent—both in providing a framework or focus for ongoing activities and for fostering new alliances and partnerships across sectors and institutions that energize schools, communities, teachers, and students.

When EHR program evaluation can contribute to the understanding of systemic reform, it will serve both the accountability and planning functions that NSF is now willing to embrace.[7] The fiscal environment notwithstanding, the stewardship of Federal funds must blend information, prudence, and vigilance to make decisions that will communicate to and benefit a host of constituencies. Only then will EHR be able to declare with confidence that we are making a difference.

Systemic Reform:
What We Expect-+

1. Improvement of All Students' Performance: "Reduce the Gap"
2. Quality Leadership at All Levels
3. Enabling Policies/Practices/Partnerships/and Organizational Structures
4. Alignment and Implementation of Quality Standards, Curriculum and Instruction, and Assessment
5. Continuous Review and Evaluation
6. Resources Reallocation

Figure 4

ENDNOTES

[1]These include presentations to EHR staff in November 1993, and subsequently to the American Association for the Advancement of Science's Science Linkages in the Community (SLIC) national advisory panel, Informal Science (an interest group composed of museum, media, and community-based organizations that bring science to the public outside of the formal system of schooling), the National Research Council's Office of Scientific and Engineering Personnel, and NSF's own Office of the Director, which reviewed EHR's programs and impacts in September 1994. This paper especially reflects the presentations made in November 1993 and September 1994. For the ideas supplied during discussions with the following EHR colleagues, I am grateful: Midge Cozzens, Joe Danek, Jim Dietz, Janice Earle, Susan Gross, Peirce Hammond, David Jenness, Con Katzenmeyer, Dick Lesh, Flo Stevens, Jane Stutsman, Larry Suter, and Luther Williams.

[2]Also see Floraline Stevens et al., *User-Friendly Handbook for Project Evaluation: Science, Mathematics, Engineering and Technology Education* (NSF93-152 new), and an accompanying video to assist project directors and evaluators in the field. In addition, the EHR staff offers workshops on evaluation at various sites throughout the United States, sometimes in conjunction with meetings of professional associations whose members include educators and educational researchers.

[3]An August 1994 EHR workshop on Developing a Dissemination Strategy addressed these very issues, which are presented in “Prospectives on Dissemination and the Dissemination Process,” an interpretive summary from the EHR/RED special invitational workshop, Division of Research, Evaluation and Dissemination for EHR/RED, August 30, 1994 (unpublished). It contains a rationalized strategy for how the Directorate can extend the reach and impact of its programs on the field.

[4]See Joseph Danek et al., "NSF's Programmatic Reform: The Catalyst for Systemic Change," in *Building the System: Making Science Education Work*, briefing book for the NSF Invitational Conference, February 24-26, 1994.

[5]SRI International, *Evaluation of the National Science Foundation's Statewide Systemic Initiative (SSI) Program: First Year Report*, Vol. 1: Technical Report, June 1994 (NSF94-95 new).

[6]This is something I learned while directing a policy study for the congressional Office of Technology Assessment, subsequently published as U.S. Congress, OTA, *Educating Scientists and Engineers — Grade School to Grad School* (U.S. Government Printing Office, June 1988) and a companion report, *Elementary and Secondary Education for Science and Engineering* (December 1988).

[7]An Office of Science and Technology Policy (OSTP) working group of which I was a member produced a draft paper, S.E. Cozzens et al., “Evaluation of Fundamental Research Programs: A Review of the Issues.” A Report of the Practitioners’ Working Group on Research Evaluation, August 15, 1994, that makes the following point: No federal program should be exempt from the kind of evaluations that policymakers need to help guide the appropriation of discretionary monies in the national interest. Evaluations cost too much in dollars, time, and energy — much like the target programs themselves — not to use as a systematic input to those decisions. Also see Daryl E. Chubin, “Meeting the Challenges of Performance Assessment.” In *AAAS Science and Technology Policy Yearbook 1994*, A.H. Teich et al., editors, the American Association for the Advancement of Science, 1994, pp. 303-305.

Appendix

DEFINITIONS OF EHR PROGRAM EVALUATIONS

March 1994

According to the Expert Panel for the Review of Federal Education Programs in Science, Mathematics, Engineering, and Technology (*Sourcebook*, August 1993, pp. 61-62), there are two categories of program review suggested for core science education programs in Federal agencies—one makes judgments of program merit and can be called program evaluation; the other collects and reviews descriptive statistics about programs and is referred to as program monitoring.

While respecting this distinction, the evaluation staff in EHR, located in the Division of Research, Evaluation and Dissemination (RED) has devised a tripartite scheme:

- **Evaluations** are systematic examinations by external or third-party evaluators working under contract to ascertain program outcomes. These may be the summative or formative variety.

- **Impact studies** are also conducted by external evaluators, in some cases blue ribbon panels. These studies yield a report on processes and outcomes that is more limited in its focus, data collection, and analysis. Impact studies will usually be more formative than summative.

- **Program monitoring** is done by program officers, with technical assistance provided by EHR/RED evaluation staff. The purpose is two-fold: to collect data on program characteristics and events on a continuous basis, and to build a culture of evaluation competence among program officers. Such assistance is intended to inform the Program Officer about the extent to which program goals and management objectives are being met. Program monitoring begins with consultation between the evaluation staff and the cognizant Program Officer and Division Director; it leads to better programs and program management.

Often, a data protocol will be designed and added to the EHR Impact Database for the initial purpose of monitoring. Eventually, it will be used in impact studies and a full evaluation of the program. The three categories constitute a continuum of oversight activities. The timetable for assigning programs to a category will be updated semiannually to reflect the progress made in the current roster of evaluations, as well as emerging needs and priorities.

EHR PROGRAM EVALUATION SUMMARY,
FY 1994 TO FY 1995 Transition

(August 1994)

Division		Program	RED Evaluation Officer
DUE	*	Undergraduate Course and Curriculum	Katzenmeyer
	c	Instrumentation and Laboratory Improvement	Dietz
	c	Collaboratives for Excellence in Teacher Preparation (**p**)	Katzenmeyer
ESIE	**c**	Informal Science Education	Sladek (Stevens)
	c	Instructional Materials Development (**p,d**)	Katzenmeyer
	c	Young Scholars (**i**)	Katzenmeyer
	c	Presidential Awards for Excellence in in Science and Mathematics Teaching (**i**)	Katzenmeyer
	c	Teacher Enhancement (**p,m**)	Katzenmeyer
GERD	**c**	Women in Engineering (**p, m, d**)	Gross
	c	Graduate and Minority Graduate Fellowships (**p, m, d**)	Gross
HRD	*	Research Improvement in Minority Institutions/Minority Research Centers for Excellence	Dietz (Stevens)
	c	Research Careers for Minority Scholars (**p, d**)	Dietz
	c	Programs for Persons with Disabilities (**needs assessment**)	Gross (Stevens)
OSR	**c**	EPSCoR (**d**)	Gross
	*	Statewide Systemic Initiatives	Gross
	c	Rural Systemic Initiatives (**m**)	Gross
RED	*	FCCSET Assessment of Federal Laboratory Capacity for Teacher Enhancement	Katzenmeyer
	c	FCCSET Evaluation of Teacher Enhancement Programs	Katzenmeyer
	c	Research in Teaching and Learning	Dietz

Key:

*	under way and continuing from FY 1993
c	current year, i.e., FY 1994, start
(**d**)	a substantial part of the evaluation is based on the EHR Impact Database
(**i**)	impact study (see Definitions)
(**m**)	program monitoring with RED providing technical assistance to the Program Officer (see Definitions)
(**p**)	planning to begin in year indicated, not actual evaluation

Interagency Efforts to Review and Evaluate Science, Mathematics, and Engineering Programs Through the Federal Coordinating Council for Science, Engineering and Technology

Joan L. Herman
Center for the Study of Evaluation
University of California-Los Angeles

James S. Dietz and Conrad G. Katzenmeyer
Division of Research, Evaluation and Dissemination
Directorate for Education and Human Resources
National Science Foundation

This article summarizes three presentations regarding the efforts of the Federal Coordinating Council for Science, Engineering and Technology (FCCSET) and its successor the National Science and Technology Council (NSTC) to promote and coordinate evaluation of Federal science, mathematics, engineering, and technology (SMET) education programs. The presentations focused on the findings of an Expert Panel organized to review Federal SMET programs and their evaluations and on subsequent steps pursued by the Evaluation Working Group that organized the Expert Panel. Joan Herman was a member of the Expert Panel. James Dietz was the principal NSF staff member providing support for the panel. Conrad Katzenmeyer continues as Co-Chair of the Evaluation Working Group.

In the transition from the Bush to the Clinton Administration, FCCSET was superceded by the National Science and Technology Council, and the parent Committee for the Evaluation Working Group became the Committee on Education and Training rather than the Committee on Education and Human Resources.

Overview

...the United States simply does not have the luxury of supporting the wrong programs or failing to support the right ones.”

-Expert Panel

The Federal Government has a substantial commitment to education in science, mathematics, engineering, and technology (SMET). At least 16 agencies provide some type of educational support in these areas. Until recently, however, each of these agencies pursued its own strategy with no coordination and often with little knowledge of what other Federal units were doing. This paper summarizes the first attempt to review agency programs and their evaluations as a whole.

In 1990, the President’s Science Advisor, D. Allan Bromley, established a Committee on Education and Human Resources (CEHR) within the Federal Coordinating Council for Science, Engineering and Technology (FCCSET). To that point FCCSET’s com-mittees had been concerned with interagency planning on research issues, such as hazardous waste, and had proven to be an effective mechanism for establishing priorities across Federal agencies. The FCCSET Committees were coordinated by the Office of Science and Technology Policy headed by the President's Science Advisor.

Under the direction of Secretary of Energy James Watkins, CEHR began its efforts by developing an inventory of SMET education programs in its participant agencies (Table 1) and compiling an aggregated budget for those programs. That was the first effort by the Federal Government to identify what it is spending on SMET education and to clarify the nature of that commitment. This effort yielded an estimate of approximately $2.5 billion in programs aimed exclusively at science and mathematics education.

The inventory and budget aggregation were reported in *By the Year 2000, First in the World*, issued in February 1991. The next task was to develop a strategic SMET education plan across the agencies. This process took over a year to complete and included the establishment of working groups addressing specific areas of science and mathematics education: elementary and secondary, undergraduate, graduate, and public understanding of science. The strategic plan, *Pathways to Excellence*, was published in January 1993, and updated in *Investing in the Future*, issued later in 1993.

In the 1992 work on the strategic plan, working groups were added on technical training, technology, and evaluation, reflecting the importance of these areas as well as the recognition that these topics cross-cut the issues considered by the other working groups. For the sake of efficiency, it was better to address the topics of technology and evaluation once rather than to try to insert them in each of the other working group reports.

The Evaluation Working Group of CEHR set the following goals for evaluation in its member agencies:

- All SMET education programs will be evaluated in a continuous, multiyear cycle;
- Agencies will be responsible for evaluating their own SMET education programs; and
- CEHR will coordinate evaluations across agencies.

To achieve these goals, the Evaluation Working Group established the following activities and outcomes.

- Under NSF leadership, the Evaluation Working Group would create an Expert Panel to inform CEHR agencies of evaluation needs. The Expert Panel would report to CEHR on the assessment of the

Table 1. CEHR Members

Department of Agriculture	Department of Commerce
Department of Defense	Department of Education
Department of Energy	Environmental Protection Agency
Department of Health and Human Services	Department of Housing and Urban Development
Department of the Interior	Department of Justice
Department of Labor	Department of Transportation
National Aeronautics and Space Administration	National Science Foundation
Smithsonian Institution	Department of Veterans Affairs
Office of Management and Budget	Office of Science and Technology Policy
National Economic Council	

merits of member agency programs and the Federal strategy.

The Expert Panel was formed in 1992 and completed its report in 1993, as discussed below.

- The Evaluation Working Group would assist in the design of an assessment study on the capacity, roles, and accessibility of Federal laboratories for teacher enhancement.

  The study was designed and carried out, with the report completed in 1995.

- Each CEHR agency would develop plans for evaluating its science, mathematics, engineering, and technology education programs. The plan would include those programs for which an evaluation would be completed by 1998 and would indicate the year(s) in which each evaluation would be conducted.

  Not all CEHR agencies completed their evaluation plans. However, a number of them did with positive results as discussed below.

The Expert Panel

In October 1992, the FCCSET Committee on Education and Human Resources chartered the planned external panel of experts to provide advice and recommendations on Federal SMET education programs and program evaluation practices. The panel—known officially as the Expert Panel for the Review of Federal Education Programs in Science, Mathematics, Engineering, and Technology—was made up of 15 experts (external to the Federal Government) representing SMET research, education research, K-12 teaching, educational administration, and program evaluation.

Recognizing the need to seek an external review of its programs, policies, and evaluation practices, FCCSET CEHR charged the panel with two tasks:

(1) To conduct a broad review of Federal programs in SMET education, and

(2) To assess Federal program evaluation efforts.

The report of the panel, *The Federal Investment in Science, Mathematics, Engineering, and Technology Education: Where Now? What Next?* presents their findings and recom-mendations. Thirteen of the 16 FCCSET CEHR member agencies participated and were the subject of the panel's inquiry, which was organized, supported, and staffed by NSF (see Appendix I for more information about how the Expert Panel was organized and Appendix II for biographies of its members). A summary of the report is given in Appendix III in the format of a set of overhead project transparencies that were presented at an EHR Evaluation Forum. Appendix IV presents select findings and recommendations of the panel—particularly concerning the organization of Federal SMET education programs and program evaluations—and provides further insight into their implications for implementation.

Federal SMET Education Programs and Budgets

The panel made two principal findings and recommendations that correspond to the panel's dual charge to broadly review the Federal programs in SMET education and to examine Federal evaluation efforts. The panel's verbatim words are in *italics.*[1] Following each set of findings and recommendations is the author's commentary (in regular type). The panel's major finding that involves the support of SMET education programs follows.

Principal Finding One

The Investment Portfolio. *The Federal commitment of dollars to SMET education is significant. In 1993 alone, $2.2 billion in Federal funds will be expended on nearly 300 "core programs" constituted solely to*

[1]This paper will focus primarily on the two principal recommendations and the recommendations made in the evaluation section of the panel’s report.

support SMET education (see figure 1). If the SMET education components of Federal "contributing programs" are included, this sum could be as large as $24.4 billion. Unfortunately, though, the Federal portfolio of core programs is unbalanced and lacks coherence. This situation is the result of varying agency missions, a decentralized congressional resource allocation process, and an overall lack of coordination and planning. The lack of coherence and balance in programs makes it next to impossible to maintain fidelity to the overarching national goals for science, mathematics, engineering, and technology education.

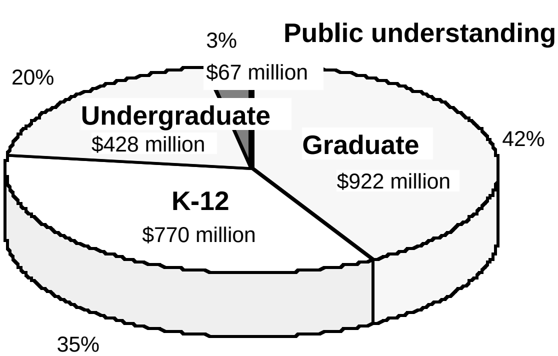
Not only is the investment significant, it is also concentrated. About 86 percent of this money is concentrated in 4 of the 11 FCCSET CEHR agencies that submitted budgets for this purpose (table 2).

Table 2. Federal SMET education budget allocation, by agency

Agency	Percent of FY 1993 budget	Cumulative percent of FY 1993 budget
NSF	25	25
DOD	24	49
HHS	21	70
ED	16	86
DOE...................	5	91
DOI....................	4	95
NASA................	4	99
USDA................	1	100
SI, EPA, DOC ...	1	101

KEY: National Science Foundation (NSF), Department of Defense (DOD), Department of Health and Human Services (HHS), Department of Education(ED), Department of Energy (DOE), Department of the Interior (DOI), National Aeronautics and Space Administration (NASA), Department of Agriculture (USDA), Smithsonian Institution (SI), Environmental Protection Agency (EPA), Department of Commerce (DOC).
NOTE: Percents may not add to 100 due to rounding.

Figure 1. FY 1993 Federal budget for SMET education, by level of education



Thus, policy or budget shifts must at a minimum concentrate on the "Big Four" agencies (NSF, DOD, HHS, and ED) in order to have some overall impact. This strategy would be particularly necessary in addressing the panel's finding that the investment is currently "unbalanced." The panel does not fully explain what it meant by this designation, but there are some clues in the budget as to what this may mean.

The Big Four agencies, mostly due to high concentrations of resources in graduate education in HHS and DOD, have allocated nearly half of their investment to graduate education and less than half of 1 percent on public understanding of SMET (table 3). On the other hand, the "other agencies"—the smaller players in SMET education—have an allocation of resources across education levels that is much more balanced than that of the Big Four agencies—the larger players in SMET education.

Table 3. Percent of budget allocation, by educational level of recipient institution and agency grouping

Agency group	K-12	Under-graduate	Graduate	Public under-standing	Total
Big Four	37	18	45	0(<.5)	100
Other agencies	25	31	25	18	99

NOTE: Percents may not add to 100 due to rounding.

By stating that the investment is unbalanced, the panel may have also meant that the budget does not reflect overall FCCSET CEHR budget priorities. If this is what they meant, they were again correct. According to the FCCSET CEHR Strategic Plan[2] that

[2]See "Pathways to Excellence: A Federal Strategy for Science, Mathematics, Engineering, and Technology Education, U.S. Science, Mathematics, Engineering, and Technology Education Strategic Plan, FY 1994-FY 1998," Committee on Education and Human Resources (Washington, D.C.: Federal Coordinating Council for Science, Engineering and Technology Committee on Education and Human Resources, January 6, 1993) or "Investing in Our Future: Science, Mathematics, Engineering, and Technology Education," Report of the FCCSET Committee on Education and Human Resources, FY 1994 Budget Summary (Washington, D.C.: Federal Coordinating Council for Science,

was examined by the panel, elementary and secondary education is the top FCCSET CEHR budget priority. From neither the Big Four nor the other agencies is K-12 receiving the largest share of the Federal investment in SMET education, and the percentage increase in the FY 1993 budget versus the FY 1992 budget did not reflect this the priority (table 4).

Table 4. FY 1992-93 Federal budget increase for SMET education, by educational level (for all agencies)

Educational level	Percent increase
Graduate	12
Undergraduate	11
K-12	9
Public understanding	3

Engineering and Technology Committee on Education and Human Resources, FY 1994 Budget Summary).

Another way of examining the question of balance is to examine the programs themselves. The panel makes reference to the fact that the $2.2 billion is spread to more than 300 programs. But how are the dollars allocated to these programs? We find that of the 290 programs identified by the panel, more than half are in the "other agencies" group that accounts for only a fraction of the budget (tables 5 and 6).

Table 5. Number of Federal SMET education programs, by agency grouping and percent of the FY 1993 budget

Agency	Number of programs	Percent of FY 1993 budget
Big Four	141	86
Other agencies	149	14
Total	290	100

Table 6. Average dollar size of SMET programs, by agency grouping and educational level (in millions)

Educational level	Big Four	Other agencies
K-12	$11.7	$1.4
Undergraduate	8.2	2.6
Graduate	24.8	2.6
Public understanding	1.4	2.9

This perhaps in part explains the FCCSET CEHR difficulty in reallocating resources; there are many programs that have very small budgets. That is not to say they are not serving a need. But the question goes beyond simply satisfying a need. Again, in the panel's words, *The lack of coherence and balance in programs makes it next to impossible to maintain fidelity to the overarching national goals for science, mathematics, engineering, and technology education*. The heart of the question is the ability of the Federal Government to <u>lead</u> the Nation toward meeting the National Education Goals or any other such goals as it may choose. Such leadership—and coordination—would seem difficult to accomplish with the large number of small programs.

To address this finding, the panel makes the following programmatic principal recommen-dation.

Principal Recommendation One

Manage the Investment: *The work of the Committee on Education and Human Resources and its Federal Strategic Plan outlined in Pathways to Excellence constitute a strong beginning--but a stronger management plan is crucial. The management plan should designate lead agencies for Federal initiatives in particular areas and recommend the merger or phasing out of existing programs, as well as the development of new programs, as appropriate. This management plan must treat Federal SMET education programs like a portfolio of investments by ensuring that a greater proportion of agency programs*

(1) are aligned with overall Strategic Plan goals,
(2) are coordinated across agencies and education levels,
(3) use effective strategies for dissemination,
(4) include appropriate evaluations, and
(5) promote equity.

Active and continuous dialogue within and among agencies (dialogue that includes state, local, and private-sector players when appropriate) must be based on a renewed commitment to effective communication and active coordination of effort.

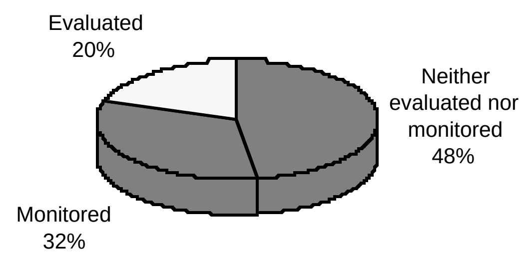
The Role of Evaluation

The panel examined existing Federal program evaluations, agency policies and other agency documents, and FCCSET CEHR planning documents concerning evaluation. After reviewing these materials, the Expert Panel concluded that evaluation of Federal SMET education programs is inadequate in terms of its quality, the speed at which it is completed, and the number of studies that have been completed. Therefore, they concluded, the effects and effectiveness of much of the Federal investment remain unexamined.

Principal Finding Two

Evaluation of the Investment. *Current SMET education evaluation practices are often inadequate for the purposes of improving programs, making informed decisions about program retention or expansion, or providing for real accountability. Funding for evaluation (FY 1993 $8 million) constitutes less than one-half of 1 percent of core Federal funding for SMET education, and in fact, just 20 percent of the approximately 300 core Federal SMET programs have been evaluated (see figure 2).*

Figure 2. Percent of Federal SMET education programs evaluated or monitored



Just one in five Federal SMET education programs have been evaluated. About 32 percent of the programs have been or are being monitored. These data were provided by the agencies, which were asked to decide what did or did not meet the definitions of evaluation versus monitoring. Had the panel categorized programs as having been evaluated or monitored, it is likely that the proportions would have been lower. The panel did not fully agree with what the agencies qualified as evaluation and monitoring. The panel wrote that, *Current SMET education evaluation practices are often inadequate for the purposes of improving programs....* The definitions of evaluation and monitoring used by the agencies and in the panel report are as follows:

Evaluation

Type A The systematic determination of merit or intrinsic worth, which includes data collection, is usually conducted by an external evaluator and examines expected and unexpected programmatic outcomes.

Type B A judgment of merit, based on existing or easily obtainable evidence, is usually conducted by an external team with a focus on expected programmatic outcomes.

Monitoring

Type C Monitoring through the collection of indicator data is usually conducted internally on a continuous basis to provide formative information about expected programmatic outcomes.

Type D Determination of the extent to which goals/management ob-jectives have been met is generally conducted internally through the use of existing data.

A higher proportion of K-12 programs have been evaluated than any other educational level category of programs. This is in concert with the FCCSET CEHR priority on K-12 education. By contrast, only 12 percent of undergraduate programs (less than half the proportion of K-12 programs) have been evaluated (table 7).

Table 7. Number of SMET education programs and percent evaluated, by educational level of programs

Educational level	Number of programs	Percent evaluated
K-12	116	26
Undergraduate	76	12
Graduate	61	18
Public understanding	25	20
Nontargeted*	12	17
Total	290	20

*Nontargeted programs are those that either defied the educational level categorization (i.e., NSF's evaluation or dissemination functions) or targeted multiple educational levels and could not be categorized (i.e., EPA's Progression Education program).

A quick review of the program evaluation policy and procedures of the FCCSET CEHR agencies reveals that few agencies are aggressively evaluating their SMET education programs, and few have organizational bodies established within the agency that are at least partly responsible for SMET education program evaluation.

- Department of Agriculture (USDA): The Secretariat of Science and Education has a congressionally funded administrative account for evaluations—the only USDA source of support other than program administration—that is often inadequate for conducting an evaluation. These funds are allocated to the agencies of USDA on a competitive basis yearly. Universities conduct these reviews for USDA by means of cooperative agreements.

- Department of Defense (DOD): A Science and Engineering Education Panel was formed in 1991 to assess DOD programs in SMET education. The panel is beginning to perform program evaluation and review in accordance with the DOD Management Plan for Science and Engineering Education. The panel will assess the effectiveness of DOD's programs and activities in meeting overall program objectives through annual reviews, the first of which was submitted in January 1993.
- Department of Education (ED): Program offices routinely gather data to monitor operations, primarily through reports from grantees and site visits by ED staff. In addition to routine monitoring, ED conducts program evaluations. A centralized unit, the Planning and Evaluation Service, administers contracts to evaluate ED's

programs. Typically, evaluations have been summative, emphasizing experimental and quasi-experimental designs. However, a broader set of approaches including case studies is now common. Although Congress often mandates that specific programs be evaluated, ED has some flexibility in selecting additional programs to be evaluated.

- Department of Energy (DOE): The coordination responsibility for DOE's university and science education activities and their evaluation lies with the Office of Science Education and Technical Information. The office supports external and internal evaluation of these programs. Other DOE units also sponsor education programs; the individual units determine how these programs will be evaluated. External evaluation for elementary and secondary programs is provided by a 4-year grant to the National Center for Improving Science Education. DOE's evaluations are funded from within individual program budgets.

- Department of Health and Human Services (HHS): Within the last several years, the Public Health Service, which supports intramural and extramural programs in life sciences education, has adopted a policy that all new programs will have an evaluation component. Each Public Health Service agency has a central planning and evaluation division that is the focal point for program evaluation and has trained evaluators on staff. Contractors are also used for evaluations. The Public Health Service Act permits the Secretary of HHS to allocate up to 1 percent of the budget for program evaluation studies. Evaluations of individual science education projects are usually supported under each grant awarded by HHS.

- Department of the Interior (DOI): There is no central DOI office that evaluates education projects; each agency is responsible for evaluating its own programs. The Bureau of Indian Affairs (which operates 183 schools) has an education evaluation unit; other bureaus have evaluation units that are not specifically geared to education. The Bureau of Indian Affairs uses agency employees and external experts to evaluate its programs.

- Environmental Protection Agency (EPA): In the past, EPA has performed a limited number of program reviews. Currently, however, newly created EPA programs are required to include plans for program monitoring and/or evaluation. EPA has no centralized evaluation unit but plans to set aside funds for a limited number of programs to be evaluated in house and by external groups.

- National Aeronautics and Space Administration (NASA): A Technology and Evaluation Branch in the Education Division was established in November 1991. The branch has agency-wide management and evaluation responsibility for education programs. NASA is developing a computer data base to store and generate reports on evaluations conducted on agency-wide programs. Evaluations are conducted internally and externally; new programs are required to include an evaluation plan before they can be approved. NASA has contracted with the National Research Council for the development of statistical indicators for evaluation.

- National Science Foundation (NSF): The Division of Research, Evaluation and Dissemination has been evaluating education programs since 1991. A staff of three plans evaluations, constructs requests for proposals for these services, oversees contractors, and provides evaluation services internally to all education and human resources programs. A formal plan to evaluate all NSF SMET programs calls for each program to be evaluated on a cyclical basis and requires that each new grant include an evaluation component.

- Smithsonian Institution (SI): The Office of Special Assistant for Institutional Studies, established in 1987, guides and assists SI units in evaluating their programs. Informal assessments and small-scale studies conducted by the individuals in charge of the programs—rather than more formal and independent reviews—characterize the evaluation of educational programs at the Smithsonian Institution.

(These summaries and the above definitions of evaluation were taken from *The Federal Investment in Science, Mathematics, Engineering, and Technology Education: Where Now? What Next? Sourcebook.)*

Big Four agencies and other agencies each monitored about one-third of their programs, although almost double the proportion of Big Four agencies' programs were evaluated compared with the programs of the other agencies (table 8). The panel identified no instances, in any agency, where a formal needs assessment had been performed before SMET education programs were created.

The panel made the following principal recommendation about evaluation.

Table 8. Percent of SMET education programs evaluated or monitored, by agency grouping		
	Evaluated	Monitored
Big Four*	26%	33%
Other agencies	14	32

*The Department of Defense failed to report that any of its programs were undergoing evaluation or monitoring.

Principal Recommendation Two

Improve the Investment: *National needs assessment should underlie program initiatives in science, mathematics, engineering, and technology education. Programs should be evaluated rigorously for effectiveness in meeting identified needs. Evaluation results should be used as a basis for planning and revising programs and should be shared with other Federal agencies. The sharing of evaluations and evaluation results among agencies prevents duplication and wasted effort, opens opportunities for collaboration across agencies, and helps to build more successful programs within agencies.*

Conclusion

The study of the Expert Panel for the Review of Federal Education Programs in Science, Mathematics, Engineering, and Technology was the first of its kind. The work represents the first time that the entire Federal SMET education portfolio was scrutinized by an independent panel of experts. The panel made some insightful findings and valuable recommendations. The process was necessary and valuable if the Federal Government is serious in its desire to help reform SMET education in the Nation.

Interagency cooperation of the type that has been exhibited by FCCSET CEHR over the past 4 years represents an important milestone in the Federal Government's approach to problem solving. Gone are the days when agencies would work independent of each other. Here are the days where societal problems are complex and interwoven. A sophisticated approach to solving these problems necessitates interagency collabora-tion and, one might argue, integration. Each agency must bring to the table its unique skills, programs, and clientele. Information about what works and does not work in education (via program evaluation and research) needs to be shared among the agencies and with their clientele. Below are some specific critiques of the FCCSET CEHR's work and accomplishments.

1. Although Federal efforts to coordinate SMET education programming and program evaluation practices have had some successes, the failure to achieve real integration is evident.

2. The authority of FCCSET CEHR to make binding decisions about the organization and operation of SMET education programs is limited. This is a basic

inability to transform what the panel report calls a “haphazard” array of programs into an integrated set of programs that can truly be called a portfolio of investments in our Nation's future. The panel strongly recommends the development of a FCCSET CEHR management plan and the designation of lead agencies in particular areas of SMET education to combat this problem.

3. Federal agencies, many of which operate SMET education programs as only a minuscule fraction of the greater agency mission, do not have the evaluation infrastructure necessary to support good management and sound decision making.

4. Agencies are torn between two education and human resources goals: to aid the reform of education (NSF, ED, and others) versus to develop a human resources base that supports the long-term ability of the agency to meet its mission (DOD, HHS, and others)

5. Perhaps the successor organization to FCCSET CEHR needs to think of agencies in terms of agency roles that could account for a diversity of agency types and sizes. One way of categorizing is based on agency budget size for SMET education (i.e., the Big Four agencies and the other agencies); another is based upon mission (i.e., SMET education and other missions).

6. Because FCCSET CEHR has failed to put forth a clear set of outcome-based goals, evaluation and its ultimate impact on programs and policies has taken on an awkward role that is outside the mainstream of decision making and policy making. Until the Federal investment in SMET education can be organized around a set of realistic goals, evaluation cannot play a central role in shaping programs and policies.

7. The Federal agency SMET education program evaluation infrastructure is weak and underdeveloped. The panel found a lack of sufficient resources and an overall lack of quality in the evaluations that have been conducted.

Implementing Evaluation Plans

A key milestone of the Evaluation Working Group is the development of evaluation plans for all the SMET programs in the CEHR member agencies. While it was possible to identify the types of programs and the amounts of money that agencies invested in them, there was little information on the impact of these programs. This was confirmed by the Expert Panel convened by the Evaluation Working Group. The panel concluded that "The Federal Government cannot continue to spend large sums of money without knowing if its programs are accomplishing their established goals—or if these goals address national needs in SMET education."

To meet the milestones delineated in *Pathways to Excellence*, the Evaluation Working Group assessed the conditions with which it was faced. Few agencies had any existing evaluation plans. In fact, the evaluation staffs of many agencies had been eliminated during the budget cuts of the early 1980s. Even where trained evaluators existed, as in HHS, evaluation of education programs was not a high priority. Only ED and NSF had existing evaluation staffs that focused on education programs. There was also little money budgeted for evaluation in most agencies. Evaluation had not been seen as a fundamental part of program efforts and was often viewed as a drag on program funds.

To address these conditions, the Evaluation Working Group proposed an evolutionary process that would prepare agency staff at the same time that evaluation plans were being developed. In most instances, money to support evaluation would accumulate gradually.

Developing Individual Agency Evaluation Plans

The first step was to identify a contractor to assist the agencies; the contractor was Westat, Inc., with Dr. Joy Frechtling as the Project Director. Westat and the Evaluation Working Group began by designing a template that would walk an agency staff member through the necessary steps to create a plan (Appendix V). Emphasis is placed on identifying the type of evaluation, the questions to be answered, the design to be employed, and the source of funds for the evaluation. Westat also built a data base to organize and standardize information from each of the agencies.

To introduce the CEHR agencies to the template, a full-day workshop was led by Dr. Frechtling and Dr. Elizabeth DeStephano of the University of Illinois-Champaign. The workshop stressed practical approaches to evaluation as well as models that might be employed.

On the basis of the workshop experience and with the assistance of Westat staff, agencies began to construct evaluation plans, a process that is still continuing. Some agencies have completed their evaluation plans while others have not, although most agencies have done some part of this task. Completed evaluation plans have been reviewed by both Dr. Frechtling and the Chair of the Evaluation Working Group and entered in the data base.

A significant finding of this work was that agencies adopted different strategies that reflect the stage of evaluation expertise and interest in the agencies. It had originally been expected that all agencies would develop a complete set of evaluation plans for their SMET education programs. For those agencies that needed to build an evaluation capacity, this was generally the case. They saw the building of evaluation plans as an opportunity to communicate with program staff and to begin legitimizing the role of evaluation.

For those agencies in which evaluation was already well established, however, develop-ment of a complete set of evaluation plans was not efficient. In these agencies, preparation of evaluation plans is a well understood task that occurs at a specified point in the process. To attempt to prepare such plans ahead of that time would be seen as a waste of program officers' time and the resulting plans would have no operational use. Therefore, the only programs for which full-scale plans were developed were those for which evaluations were due.

Developing a Master Plan

The second part of the development of agency evaluation plans was to prepare a master plan across the agencies. This master plan would feature joint efforts on program topics of high mutual concern.

Plans for pursuing joint evaluation efforts across agencies have proceeded. It is clear that the most visible set of programs in the FCCSET agencies are those addressing teacher development or enhancement. The largest Federal SMET education program, ED's Eisenhower Program, is exclusively involved with teacher development. NSF has traditionally supported teacher institutes and continues to support a major program in this area. The mission agencies, such as DOE and NASA, have substantial teacher projects in their facilities. Given that CEHR set as its goal to provide intensive disciplinary and pedagogical training to 600,000 teachers by 1998, the need for evaluations of the efficacy of these programs is obvious.

The teacher development/enhancement evaluation began in the summer of 1994. The focus is on developing a summative, indepth evaluation of these programs, as most previous evaluations have tended to emphasize the number reached rather than efficacy. The initial phase emphasized program identification and site visits to the programs during the summer development program. The second phase focused on determining the impact of these programs in the teachers' schools and classrooms through surveys and case studies.

The study has confirmed that there is a professional agreement on what is best practice in teacher development projects. It is also clear that the strongest of the Federal teacher development projects are effective in creating a hands-on science environment for their participants. Creating conditions supportive of systemic reform in these projects is less certain; agencies are just learning how to build systemic reform into teacher develop-ment/enhancement.

Other topics of mutual evaluation interest are also being pursued. A major area of concern among the agencies is the design and impact of programs to serve those groups underrepresented in mathematics and science. An evaluation of these programs is underway, as is specification of criteria for all SMET programs that address increasing participation of underrepresented groups.

Evaluation training is another area of collaboration. There is particular interest in workshops on developing and measuring performance indicators for SMET programs. A joint effort is underway to specify educational indicators for these programs and training agency personnel on their use.

Conclusion

FCCSET's CEHR process has been an example of agency staff attempting to achieve a set of goals without having a clear mandate from many of the agencies. This has been particularly true in evaluation. Although the agencies signed off on the strategic plan by saying they were going to evaluate all of their programs, in fact few provided any money, and almost none added any specialized staff.

Some have decried the slow pace and the sometimes weak actions that have resulted. However, in times of decreasing budgets and shrinking staffs in many agencies, this may be the most that can be hoped for. FCCSET and CEHR have achieved a good deal of visibility and have created conditions that have led to individuals in various agencies working effectively together. This probably will not have major impacts on agency programs, but it has the potential to provide a beginning for evaluation in many agencies that was not there before.

References

By the Year 2000: First in the World. Report of the FCCSET Committee on Education and Human Resources, February 1991.

By the Year 2000: First in the World. Report of the FCCSET Committee on Education and Human Resources, January 1992.

Pathways to Excellence: A Federal Strategy for Science, Mathematics, Engineering, and Technology Education. Report of the FCCSET Committee on Education and Human Resources, 1993.

Investing in Our Future: Science, Mathematics, Engineering, and Technology Education. Report of the FCCSET Committee on Education and Human Resources, 1993.

The Federal Investment in Science, Mathematics, Engineering, and Technology Education: Where Now? What Next? Report of the Expert Panel for the Review of the Federal Education Programs in Science, Mathematics, Engineering, and Technology, June 1993.

The Federal Investment in Science, Mathematics, Engineering, and Technology Education: Where Now? What Next? Report of the Expert Panel Sourcebook, August 1993.

Appendix I

Expert Panel for the Review of Federal Education Programs in Science, Mathematics, Engineering and Technology

The Panel Process

The Expert Panel met three times between October 1992 and March 1993 to deliberate on findings and recommendations. The meetings provided the opportunity to discuss issues and to meet with agency representa-tives to clarify understanding of each agency's programs. Panelists focused their examination on Federal SMET education core programs[1] (these are programs that operate for the express purpose of improving SMET education). Much of the work of the individual panel members was conducted off-site in the periods between meetings.

The panel was organized into three five-member subpanels, each responsible for examining the programs and program evaluations for a select group of agencies. These subpanels were chaired by the three evaluators on the panel. In addition, each panel member was assigned to one of five topical areas. Five topical groups of three members each (one member from each subpanel) covered elementary and secondary education, undergraduate education, graduate education, public understanding of science, and program evaluation. This matrix structure ensured that both agencies and educational areas of interest would be covered during the panel's deliberations. The panel conducted its deliberations through three 2-day meetings, with work continuing in the intervening time.

The panel used existing materials provided by the agencies and FCCSET CEHR.[2] Those materials included the following:

- Relevant plans and strategies provided by each agency and FCCSET CEHR,
- One-page statements on each program,
- A matrix of programs, budgets, and audiences affected,
- Written and oral briefings by each agency and FCCSET CEHR representatives,
- A matrix and supporting narrative on the evaluation projects of each agency,
- Evaluation reports and program audits,
- Other publications, reports, and guides,
- Curriculum materials,
- Sample surveys,
- Program guides and inventories, and
- Information about the condition of SMET education in the United States.

The panel was co-chaired by Karl S. Pister, an engineer and Chancellor of the University of California, Santa Cruz, and Mary Budd Rowe, Professor of Science Education, Stanford University (see Appendix II for biographies). Finally, the report from the panel is wholly their own. The findings and recommenda-tions contained in their report and in this document represent the views of the panel, not the National Science Foundation, the FCCSET

[1] The panel members also considered Federal support for programs contributing to SMET education although not explicitly designated or managed as such.

[2] For a bibliography of selected reports and materials examined by the panel, see "The Federal Investment in Science, Mathematics, Engineering, and Technology Education: Where Now? What Next?"

CEHR, or any of the agencies participating in the study.

Appendix II

Expert Panel for the Review of Federal Education Programs in Science, Mathematics, Engineering, and Technology

Biographies

Karl Stark Pister, Co-Chair, Expert Panel; Chancellor, University of California, Santa Cruz, and formerly Dean of the College of Engineering, University of California, Berkeley, is chairman of the Board on Engineering Education for the National Research Council. Dr. Pister is a member of the National Academy of Engineering.

Mary Budd Rowe, Co-Chair, Expert Panel, and Professor of Science Education, Stanford University, is past president of the National Science Teachers Association and formerly a chairperson for the American Association for the Advancement of Science (AAAS). She is now serving on the Council and the Committee of Council Affairs of AAAS. Dr. Rowe produced the first CD-ROM available for science education, Science Helper K-8.

Stephen C. Blume, Elementary Science Specialist, St. Tammany Parish Public Schools, Slidell, Louisiana, is past president of the Society of Elementary Presidential Awardees and author and co-author of elementary and middle school science textbooks and curricular materials. He was a recipient of the National Presidential Award for Excellence in Science Teaching, 1990.

Patricia Chavez, Statewide Executive Director, New Mexico Mathematics, Engineering, Science Achievement (NM/ MESA, Inc.), is responsible for overall administration and advancement of New Mexico's successful precollege mathematics, engineering, and science achievement program. She is also the National Vice President for the National Association of Precollege Directors (NAPD), a member of the American Association for the Advancement of Science, and a member of the Mathematical Science Education Board.

Ronald L. Graham, Adjunct Director of Research at AT&T Bell Laboratories, is one of the world's leading combinatorial mathematicians. He is President of the American Mathematical Society and Professor of Mathematical Sciences at Rutgers University.

Joan L. Herman, Associate Director, National Center for Research on Evaluation, Standards and Student Testing at UCLA's Graduate School of Education, is the author of *Tracking Success: A Guide for School-Based Evaluation* and the editor *of Making Schools Work for Underachieving Minority Students.*

Ernest Robert House, Professor of Education and Director of the Laboratory for Policy Studies at the University of Colorado, Boulder, is the author of *Professional Evaluation Social Impact and Political Consequences*. He is the winner of the Lazersfeld Award for Evaluation Theory in 1990 and the Harold D. Laswell Prize awarded by Policy Sciences in 1989.

Jacquelyn S. Joyner, Mathematics Instructional Specialist K-12, Richmond, Virginia, Public Schools, served as a member of the National Advisory Board, Macmillan/McGraw Hill, and was commissioned by the National Center for Education Statistics to write a paper on the mathematics items of the National Assessment of Education Progress (NAEP) examination for 1991.

Floretta Dukes McKenzie, President of The McKenzie Group, a comprehensive education consulting firm, was formerly superintendent and Chief State School Officer for the District of Columbia Public Schools. In the spring of 1990 and

1991, Dr. McKenzie was a distinguished visiting professor at Harvard University's Graduate School of Education. She is presently Distinguished Urban Educator-in-Residence at The American University, Washington, D.C.

Jose Mestre, Professor of Physics at the University of Massachusetts, Amherst, specializes in cognitive processes pertaining to learning science and mathematics and is co-author of *Academic Preparation in Science*. He has served as chair of the College Board's Sciences Advisory Committee and on various national boards, such as the National Research Council's Mathematical Sciences Education Board.

Wendell G. Mohling, Teacher and Outdoor Laboratory Director at Shawnee Mission Northwest High School, Kansas, is the NASA Space Ambassador from Kansas and a member of the International Faculty for the Challenger Center. He was the 1992-93 President of the National Science Teachers Association and is a former director of the National Science Teachers Association High School Division.

Michael James Padilla, Chair, Department of Science Education, University of Georgia, is an author of numerous articles and books on science teacher education. He has been appointed by the National Science Teachers Association to various boards of the National Council for the Accreditation of Teacher Education and currently leads the Georgia Statewide Systemic Initiative.

Helen R. Quinn, Senior Staff Scientist and Assistant to the Director for Education and Public Outreach, Stanford Linear Accelerator Center, is a fellow of the American Physical Society (APS), has served on the APS Panel on Public Affairs, and is President of the nonprofit Contemporary Physics Project.

Michael Scriven, Consulting Professor, Stanford University Graduate School of Education, is a Senior Fellow sponsored by the American Educational Research Association and the National Science Foundation. He publishes, teaches, and provides consultation in a broad range of disciplines in both the physical and the social sciences. He is the editor and author of numerous publications including *Evaluation Models* and *The Evaluation Thesaurus*.

James G. Wingate, Vice President for Programs, North Carolina Community Colleges, is co-author of *Fundamentals of Probability* and has been actively involved in the Fund for the Improvement of Post Secondary Education, the National Association for Institutional Research, and the American Association of Community and Junior Colleges.

Frances Lawrenz, Special Assistant to the Panel Co-Chairs and Professor, Director of Graduate Studies in the Department of Curriculum and Instruction at the University of Minnesota, is the author of many articles on science education. She has conducted numerous evaluations of science programs. She served twice as a visiting scientist for program evaluation at the National Science Foundation.

Appendix III

Summary of Expert Panel Report[1]

The Federal Investment in Science, Mathematics, Engineering, and Technology Education: Where Now? What Next?

CONTEXT FOR EXPERT PANEL:

"...educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a nation and as a people."

A Nation At Risk, 1983

"The nation that dramatically and boldly led the world into the age of technology is failing to provide its own children with the intellectual tools needed for the Twenty-First Century."

Educating America for the 21st Century, 1983

THE EXPERT PANEL:
A CONGRESSIONAL CHARGE

- Representatives of SMET education, evaluation
- *What now*: Review of Federal Investment
- *What now*: Status of evaluation
- *What next*: How can Federal Government best help our nation achieve and maintain leadership in SMET?

MAJOR FINDINGS OF
THE EXPERT PANEL

1. Significant Federal Investment in SMET education:
    - <u>$2.2 billion</u> for programs directed solely at SMET education
    - Across 13 agencies
    - Not counting programs for which SMET goals are only a part
    - With "contributing programs," Investment about <u>$24.4 billion</u>

HOW IS THE CURRENT
INVESTMENT ALLOCATED?

Grouped according to components of Federal strategic plan:

- Pre-K-12 SMET education: 25%
- Undergraduate SMET education: 20%
- Graduate education: 42%
- Public understanding: 3%

[1] This appendix is composed of selected overheads from a presentation made by Joan Herman to NSF on December 7, 1993.

MAJOR FINDINGS OF THE EXPERT PANEL

2. The Federal portfolio is unbalanced and lacks coherence
   - *Ad hoc development*: varying agency missions, decentralized resource allocation, lack of overall coordination
   - *Fidelity to overarching Federal SMET goals* almost impossible to maintain

MAJOR FINDINGS OF THE EXPERT PANEL

3. Evaluation practices inadequate
   - Evaluation funding ($8m) is less than .5% of core funding for SMET education
   - Only 20% of 300 core programs have been evaluated
   - What counts as evaluation is problematic

PRINCIPAL RECOMMENDATIONS OF EXPERT PANEL:

1. Manage the investment: A stronger, overall management plan is crucial
   - Designate lead agencies for specific areas of SMET goals
   - Consolidate, phase out, develop new programs as appropriate
   - Promote active, continuous dialogue
   - Assure active coordination of effort

MANAGE FEDERAL PORTFOLIO, ASSURING PROGRAM INVESTMENTS:

1. Are *aligned* with overall strategic plan goals
2. Are coordinated across agencies and education levels
3. Use effective *dissemination* strategies
4. Include appropriate *evaluations*
5. Promote *equity*

PRINCIPAL RECOMMENDATIONS OF EXPERT PANEL:

2. Use evaluation to improve the investment
   - Program initiatives based on national needs assessment
   - Program effectiveness rigorously evaluated
   - Data-based planning and decision making
   - Benefits of sharing results

THE ROLE OF EVALUATION

- Evaluation is essential in sound management.
- Evaluation of Federal SMET programs inadequate.
- Effects and effectiveness of the large Federal investment is largely unexamined.
- Lack rational basis for federal strategic planning and decision making

THE EVALUATION CHALLENGE

Programs are diverse, complex, and often of great scope: Rigorous evaluation is a challenge

Agencies lack evaluation expertise

Insufficient resources are allocated: time, $$, staff

Time and commitment to use results

EVALUATION FINDINGS & RECOMMENDATIONS:

While indicators are not a substitute for comprehensive evaluation, they can play an important role in monitoring and in creating a culture which values evaluation and focuses on outcomes. There is no agreed- upon set of indicators across programs and agencies:

- Evaluation designs across agencies should include a *minimum* set of core indicators
- Indicators should be augmented by systematic studies
- Don't overburden locals, encourage use

PANEL CONCLUSION

- Enormous energy and commitment
- Many positive efforts are making a difference
- Much remains to be done
  - Strengthen and redefine Federal role in SMET education
  - New culture of coordination and communication across agencies
  - Leverage, provide leadership
  - Evaluate, promote optimal practice and sound decision making

Appendix IV

Expert Panel for the Review of Federal Education Programs in Science, Mathematics, Engineering, and Technology

Select Findings and Recommendations on Evaluation of Federal SMET Education Programs

In addition to the principal findings and recommendations, the Expert Panel made a host of specific recommendations, some of which apply directly to improving the quality of Federal SMET education evaluation practices. In its report, the panel made clear its view that evaluation is an essential component of good program management, decision making, strategic planning, and the optimal allocation of scarce resources. The panel's findings and recommendations are in *italics*. The author's commentary appears in regular type.

Evaluation Findings and Recommendations:

F1: The recently adopted Federal Strategic Plan for SMET education contains several vital program evaluation features:

- *Evaluations will be conducted in a continuous, multiyear cycle.*
- *Agencies will build the appropriate capacity to monitor evaluations.*
- *Evaluations will be coordinated and synthesized across agencies.*
- *An expert panel will advise the agencies.*

R1: The evaluation component of the Federal Strategic Plan for SMET education must be implemented. The Evaluation Working Group of the interagency Committee on Education and Human Resources must continue to monitor and more actively coordinate evaluation of SMET education programs throughout all agencies.

This finding refers to the evaluation section of the Federal Strategic Plan called *Pathways to Excellence: A Federal Strategy for Science, Mathematics, Engineering, and Technology Education*. The panel endorses the evaluation component of the Strategic Plan but is most concerned that it be fully and rapidly implemented.

Since the Expert Panel examined the Strategic Plan, FCCSET CEHR has begun to develop a Federal Master Plan for evaluation (along with individual agency plans). This Master Plan is expected to spell out just how these goals set forth in the Strategic Plan will be carried out. Much of the Federal Government's commit-ment to evaluating its SMET education programs rides on the success of this plan. However, as discussed above, few agencies are currently organized to ensure that these goals can be achieved.

In an unrelated development, the Congress has enacted the Government Performance and Results Act, which mandates that by the year 2000 all agencies' programs will be subject to a process in which programmatic goals are set and outcomes measured, documented, and compared with those goals. This process has already begun with a number of pilot projects throughout many agencies.

F2: Current efforts to coordinate evaluation activities across agencies are progressing too slowly. More attention must be given to developing cost-effective evaluations as well as an interagency capability to conduct major evaluation initiatives.

R2(a): All agencies should show evidence of significant progress in planning and implementing evaluations by the end of fiscal year 1994.

R2(b): Evaluation efforts must be prioritized and combined across agencies to not only make the most efficient use of existing funds but also allow examination of the whole Federal portfolio in terms of progress, balance, and responsiveness to changing needs. It may be possible to develop template or prototype evaluation designs that would streamline some of the evaluation process. Agencies with expertise in evaluating particular types of programs should be designated to take the lead in developing common evaluation designs.

This evaluation finding, having to do with building cost-effective techniques and interagency capabilities, has only been partially addressed. FCCSET CEHR has not begun to develop or employ newly designed cost-effective evaluation techniques.

On the other hand, under the auspices of the Federal Master Plan for evaluation, planning work has begun on interagency evaluation. Interagency evaluation means that all similar Federal SMET education programs (e.g., many agencies have teacher enhancement programs) would be evaluated under one joint study. This technique would be particularly useful in identifying those programmatic approaches that have been most effective. Or, to state it differently, the best and worst features of a particular grouping of like programs could be identified and shared among the agencies in order to strengthen all programs of that type simultaneously. Important to this effort would be to disseminate those findings to education policy makers and practitioners.

The difficulty with interagency evaluation primarily lies in how it is financed. Because of the complex rules governing Federal contracting procedures and scarce resources, interagency evaluation is easier said than done.

Recommendation 2(a) states that all agencies should show significant progress in planning and implementing evaluation by the end of FY 1994. Although some small progress has been made in interagency program evaluation planning, and some agencies have progressed further with their own evaluation activities than others, on the whole, little evidence can be found to demonstrate that *significant progress* has been made. More or less the same can be reported on R2(b).

F3: The quality, extent, and timeliness of evaluation practice vary substantially. Although evaluation design obviously depends in part on the nature of the programs themselves, and although no single set of methodologies or techniques will be appropriate for every type of evaluation, agencies and programs must nevertheless meet standards of good evaluation practice.

R3: Federal agencies should implement standards of evaluation practice, using as a base those standards currently being revised by the Joint Committee on Standards for Educational Evaluation, a coalition of 15 professional organizations concerned with the quality of evaluations.[1] Several concerns are especially relevant with regard to the setting of such standards.

- *Evaluations should be designed to minimize demands on project participants. Strategies that require all participants or all recipients to respond to extensive data collection procedures should be minimized.*
- *Timeliness is essential for evaluation studies whose results are expected to inform Government policy makers. This fact requires that current governmental clearance processes be accelerated.*
- *Information on costs and cost comparisons is critical to sound evaluation. A cost-benefit perspective should be maintained both for programs and for evaluations.*
- *Evaluations should be designed with appropriate attention to the needs assessment that justifies the program.*

---

[1]See *Standards for Evaluations of Educational Programs, Projects, and Materials*, the Joint Committee on Standards for Educational Evaluations, New York, NY: McGraw-Hill, 1981.

FCCSET CEHR has begun to address the issue of standards of good evaluation practice. A workshop has been conducted on the Joint Committee Standards and how they might be implemented.

F4: Because programs and the influences on them are complex, evaluations must examine the nature of the programs themselves as well as all intended and unplanned outcomes over extended periods of time. This is true for evaluations at all levels, from local projects to projects that cut across Federal agencies. At present, Federal agencies lack a systematic perspective on evaluation that would allow them to revise programs on the basis of assumptions, evidence of redundancies or gaps, or the clarification and validation of effective models.

R4: Evaluations within and across programs should be based on a systems view, a view that considers key factors and influences on program operation and on short- and long-term outcomes. Furthermore, evaluations should encourage the identification and dissemination of exemplary practices and should provide those who implement programs with information to help them upgrade their programs.

F5: Time, staff, expertise, and funds are inadequately allocated to the evaluation tasks at hand. Good evaluation requires a generous yet judicious commitment of resources.

R5: Funds for evaluation should be priority budget items for Federal agencies and for the projects they support. Additionally, time for learning about how to conduct evaluations and for reviewing, synthesizing, and implementing evalua-tion results should be made available to Federal agency staff.

Of course, scarce resources are often a problem in program evaluation. The key to this finding and recommendation, however, is its relationship with the principal recommendation of the panel that Federal programs be viewed and managed as a portfolio of investments in our Nation's future. This suggests that program evaluation is a necessary ingredient in operating good programs and maximizing the effect of those programs on the reform of education. This cannot be accomplished if there exists no infrastructure in several of the agencies to begin even limited evaluation projects.

F6: Many Federal agencies currently collect "indicators" to monitor program operations. Indicators are statistics about programs and their impacts; as such, they do not substitute for proper evaluation. However, indicators do play a role in program monitoring. They also aid in developing a culture of evaluation that focuses on high-priority outcomes and means of attaining them. Unfortunately, there is no agreed-upon set of indicators across agencies; each agency has its own way of collecting statistics.

R6(a): Evaluation designs across agencies should include a minimum core set of indicators to be collected and synthesized (in conjunction with other information) by program managers for similar types of programs.

R6(b): When indicators are used, they must be augmented by objective, systematic evaluation studies.

R6(c): Federal data collection efforts must not overburden local programs but must encourage local programs to use the collected information for program decision making.

FCCSET CEHR has made little progress in identifying a minimum core set of indicators that can be collected on all similar programs as recommended in R6(a). The panel made clear, however, that it believes that indicators by themselves are not enough, and they are not a substitute for thorough program evaluation.

Appendix V

Template for Creating an Individual Agency Evaluation Plan

Program and Evaluation Description

1. Name of agency

2. Name of program

3. Program area, sub-area, and targeted educational level (use categories shown in Table 1)

4. Program description

   a. Major program activity (for example, increasing teachers' awareness and familiarity with new methods and materials for the teaching of science in elementary grades).

   b. Purpose and anticipated results (for example, to increase students' understanding of scientific concepts and methods, increase their interest and competence, and to motivate these students to study science in secondary school).

   c. Specific program operations (for example, several year-long teacher workshops and summer programs at a local university; also seminars for elementary school principals).

   d. Target audience(s) and number of program participants served in each audience (for example, in 1994 a total of 1,400 teachers attended year-round workshops, 500 teachers participated in summer programs, and a seminar was held for 50 elementary school principals).

   e. Budget for FY 1993 and earlier, and for FY 1994, 1995, 1996, 1997, 1998.

5. Status of program evaluation (has evaluation been planned and/or is it currently in progress? If none planned, explain, then skip the remaining items).

PLEASE NOTE: ITEMS 6, 7, 8, 9 AND 10 SHOULD BE ANSWERED FOR ALL EVALUATIONS CURRENTLY IN PROGRESS AND THOSE TO BE INITIATED IN FY 1994. IF SOME PLANNING HAS ALREADY BEEN DONE FOR EVALUATIONS TO BE STARTED IN LATER FISCAL YEARS, PLEASE PROVIDE ANY AVAILABLE INFORMATION.

6. Performers

   a. Who will be responsible for overseeing and managing the evaluation?

   b. Will the evaluation be performed by in-house personnel only, or will contractors and/or consultants be used? What will be the respective roles of agency and outside personnel?

   c. If “in-house personnel” will perform all or part of the evaluation task, please specify units responsible for the evaluation and amount of time (FTE) required.

   d. If outside contractors or consultants are used, please describe responsibilities of each outside evaluator group, or where relevant, individuals.

7. Description of evaluation methodology

   a. What does the agency hope to learn from this evaluation? What are the specific questions which the evaluation is designed to answer?

   b. Indicate how you set priorities among the research questions. How do the questions address the needs of stakeholders?[1]

   c. Describe the evaluation design.

   d. Describe data sources. Indicate if existing data will be used (for example, project statistics, or student grades). If new data are being developed, describe types of data to be collected (for example, teacher or student reaction to the program activity, classroom observations).

[1] A program’s stakeholders are individuals or groups who may affect or be affected by program evaluation.

e. For each data collection method for newly collected data (such as case studies, indicators, surveys, expert opinions), describe in detail the methods which will be used (for example, for case studies the number of studies, how selected, what types of information will be collected; for surveys, sampling methods, data collection mode such as telephone interviews, mail questionnaires etc acceptable response rates etc.).

f. Planned data analysis (this item applies to existing data as well as newly collected data). How will data be analyzed? Describe specific quantitative and qualitative methods to be used.

8. Evaluation budget. For each year during which the program was and/or is active, show total budget and funding source.

Year	Evaluation budget	Source of funds
1993 and earlier		
1994		
1995		
1996		
1997		
1998		

9. What is the time-table for completion of the evaluation?

	Year	Month
Evaluation start-up date		
Interim report date		
Evaluation completion date		

10. Describe the reports or other products which will result from this evaluation effort, when these products will be available, and the methods you plan to use for disseminating the findings.

Fostering Change in Science Education

James D. Ellis
Senior Staff Associate
Biological Sciences Curriculum Study

The past decade has been an era of reform in science education. In the United States during the 1980s, various groups produced numerous reports denigrating the current state of education and calling for major reforms. Since 1985, educational leaders have initiated many projects to improve curriculum, instruction, and assessment in science (AAAS 1989; BSCS 1989; Bybee et al. 1990; Loucks-Horsley et al. 1990; NCISE 1991; U.S. Department of Education 1991; and NRC 1993).

Educational reform is not new. In the United States, the 1960s also was an era of reform. Past reforms, however, have failed to leave their mark on education. The changes were ephemeral at best. The central question in educational reform is, How is educational change made and sustained? This paper addresses that question as it concerns science education.

The Reform Movement

Current reform projects are redefining the why, who, what, and how of precollege education in science, technology, and mathematics. Those questions are discussed in the next sections.

Why Reform?

The *why* is the motivating force behind the reform movement. Reports from business, industry, government, and the scientific and engineering communities have decried the failure of schools to educate the Nation's work force, which has contributed to the decline in economic growth (Carnegie Commission 1991; Hurd 1986; NCEE 1983; U.S. Department of Labor 1991; Education Commission of the States 1983). On recent international assessments, the U.S. compared poorly with other countries in student achievement in science and mathematics (Lapointe, Mead, and Phillips 1988; Mullis and Jenkins 1988). Other studies indicate that to be competitive today, business and industry require a work force with improved critical thinking skills and substantial knowledge in science, mathematics, and technology (defined both as *knowledge about* technology and *use of* advanced technologies) (OTA 1988). In response, the state governors and President Bush declared war on educational mediocrity, establishing the goal, "By the year 2000, U.S. students will be first in the world in mathematics and science achievement" (U.S. Department of Education 1991).

Who Should Reform?

The *who* of the current reform movement embodies a major shift from past reform efforts. The primary focus during 1960-80 was on expanding *the pipeline* for the production of scientists and engineers. This focus was in response to a perceived national crisis during the cold war to win the space race and to be the leader in military technology. Science education targeted those students who would pursue science and technology in postsecondary institutions. In contrast, the target of the current educational reform is *science for all*. Demographics suggest the necessity of expanding the target audience for education in science, technology, and mathematics to meet adequately the projected needs of business and industry to support continued economic growth (Vetter 1988). Recent reform efforts, consequently, emphasize traditionally underrepresented, underserved populations (women, minorities, and the physically disabled) to meet the need for general scientific literacy.

What Is the Focus of the Reform?

The *what* also is changing during the current reform. The current reform movement calls for *systemic change*—reform of all components of the educational system, including curriculum, instruction, assessment, educational technology, teacher education, school organization and administration, instructional support systems, and school culture. As has been said about altering biological systems, "You can't do just one thing" (Hardin 1968). A change in one component may lead to unplanned and undesirable changes in other components. Components (i.e., parents, administrators, and school culture) in a stable, dynamic system resist and reject changes to other components in the system (such as the curriculum). To reform the system, one must address all components of the system simultaneously.

The prevailing educational system is based on the *industrial model* common in the early 20th century. The teachers are the skilled workers dispensing knowledge, the administrators make and monitor the decisions about curriculum and instruction, and the students are the products. This industrial model is consistent with the view of learning as the acquisition of information. The major goal of an industrial system is to produce a product as efficiently and effectively as possible while resisting changes that challenge the extant system. In contrast, the emergent vision of education, developed initially in business and industry, is based on the metaphor of a *learning community* in which the students are the workers, the teachers are the facilitators of learning, the administrators are instructional leaders, and the product is the knowledge coconstructed by the learners (Fullan 1993; Marshall 1990). This learning community model is consistent with the view of learning as an active construction of personal and shared knowledge.

The issue of *what* to change, therefore, rests with the idea that the education system needs to be restructured to fit this new model of a learning community. The learning community, however, must incorporate all stakeholders, not just the students. Teachers as well as students need to become lifelong learners. That is what is meant by calls for the professionalization of teaching. Science teachers first should become expert learners of science; only in that way can they become mentors for students engaged in the activity of learning science. Science teachers second should be active, lifelong students of science teaching. That is what is meant by calls for teachers that are reflective practitioners. The problem of restructuring the educational system initially is how to break through the natural, long-standing impediments to change inherent in the hierarchical, industrial model and to build a new system that fosters a culture of continual change and growth within the structure of a supportive learning community.

How to Succeed at Reform?

A new approach to *how* to reform science, technology, and mathematics education is underway. Educational leaders recognize two factors as being critical to successful reform: (1) reform requires support, commitment, and participation of all stakeholders such as teachers, administrators, college faculty, parents, business and industry, and students; and (2) reform requires a long-term commitment of material and human resources. Successful reforms are not *top-down* quick fixes to problems nor are they *bottom-up* solutions to immediate needs; they are collaborative, local programs of long-term change. National standards, state guidelines, science curricula, educational research, and assessment programs provide road maps and tools for reforming education. Changes to school programs, however, are made by teachers in local classrooms to accommo-date the unique mix of students, parents, and teachers.

Educational change takes time. A total rethinking, redesign, and reform of education may take decades. Indeed, educational leaders are begin-ning to realize that reform is a continuous process. The most productive focus for reform is on the process rather than the product, because the product is constantly changing in response to changes in society and no one static product meets the needs of a dynamic system. Enacting a culture embodying a continual process of change will allow the education system to be proactive and adaptive rather than reactive. Successful reform requires teachers, schools, states, and nations to accept the responsibility to continually assess, adapt, revise, and construct innovative approaches to science, mathematics, and technology education to serve the common good.

Recommendations for Reform

Successful reform is systemic; it simultaneously addresses all interdependent components of the educational system—the curriculum, teacher education, the instructional support system, and the school culture. Through analysis of past successes and failures and through study of the reform process and school cultures, educational researchers have uncovered key components of successful systemic reform efforts. Educational leaders can use this knowledge about the reform process to successfully implement changes in science and technology education. Successful educational reforms accomplish the following in concert:

- Coordinate all aspects of the educational program;

- Provide for the professional development of teachers;
- Restructure educational institutions to be supportive of continual change; and
- Construct a school culture promotive of educational reform.

Coordination of the Program

Curriculum, instruction, and assessment are three major components of educational programs. The *curriculum* defines the course of study, including the goals and objectives, subject matter, and specific learning activities. *Instruction* is what teachers do—the specific artful (and perhaps research-based) classroom interactions planned, initiated, and facilitated by the teacher to promote student learning. Assessment is the process by which students, teachers, administrators, and bureaucrats collect information about student learning and program effectiveness. *Assessment* provides the feedback loop in the educational system for maintaining and improving curriculum and instruction. In theory, these three major components of the school program mesh to achieve society's educational goals and aspirations. In practice, unfortunately, most current science curricula, instructional approaches, and assessment strategies are inadequate to achieve society's aspirations for a universal scientifically and technologically literate citizenry.

To successfully reform science and technology education, educational leaders must coordinate changes in the three major components of the educational program. Revisions to one component (such as changing the curriculum) are ineffective, and perhaps harmful, if concomitant changes are not made to the other components, and educational reformers should ensure that changes to the three components are based on a unifying, consistent philosophy of education. Effective curriculum developers produce materials that embody compatible recommendations for reform in curriculum, instruction, and assessment. Leaders seeking to improve the use of advanced technologies would improve success by coordinating the use of technology with general approaches to curriculum, instruction, and assessment embodied in the contemporary reform movement. Successful educational change agents (university science and education faculty and school administrators) design and conduct reform projects that coordinate improvements to all program components. Effective teachers develop an overriding philosophy that guides their approach to curriculum, instruction, and assessment.

Curriculum. Teachers and curriculum developers organize curricula in a variety of ways. Most curricula center on a single science discipline with the emphasis placed on *covering the book*. This type of curriculum stresses covering the major facts and information of a scientific discipline. Current efforts at science education reform, however, recommend a science-technology-society (STS) theme, an integrated approach, or a thematic approach to organizing science curricula. The National Science Education Standards (NSES) (NRC 1993) organize science curricula around four major themes: (1) science subject matter, (2) inquiry, (3) connections to other disciplines, and (4) science and human affairs.

Most science curricula are based on conceptions of what is worth knowing in science developed during the 1960s and earlier. Current curriculum design studies (AAAS 1989; BSCS 1989; Bybee et al. 1990; Loucks-Horsley et al. 1990; NCISE 1991; NRC 1993) call for major changes in *science subject matter*. The slogans *less is more* and *less breadth and more depth* emphasize the need for students who have meaningful understanding of science concepts that can be applied in making decisions as citizens in a global society and in solving problems in an increasingly scientific and technologic work place.

In *Science for All Americans* (AAAS 1989), AAAS provides an indepth examination of content for precollege education in science, technology, and mathematics. The fundamental premise of AAAS is, "Schools do not need to be asked to teach more and more content, but rather to focus on what is essential to scientific literacy and to teach it more effectively." AAAS brought together leading scientists and science educators to delineate the core content for scientific literacy. The major departures of *Science for All Americans* from past declarations of appropriate science content are (1) the boundaries between traditional subject matter categories are softened and connections are emphasized, (2) the amount of detail that students are expected to retain is considerably less than in traditional science, mathematics, and technology courses, and (3) the recommendations include topics not typically included in school curricula, such as the nature and history of science and technology.

The NSES suggest several approaches to subject matter, including a thematic approach. In a thematic approach, the curriculum is based on major conceptual themes of science. The National Center for Improving Science Education (Bybee et al. 1990) lists the following major conceptual themes for science: (1) cause and effect, (2) change and conservation, (3) diversity and variation, (4) energy and matter, (5) evolution and equilibrium, and (6) models and theories. For example, a unit on equilibrium might look at dynamic equilibrium in systems, human body systems, and steady-state conditions. The units are designed to help students construct personal understandings of the themes. The activities may engage students in answering a question or solving a problem and often may transcend disciplinary boundaries.

Scientific inquiry will have a prominent place in the NSES. The NSES will propose the incorporation of several perspectives of scientific inquiry in school science programs: inquiry as subject matter, inquiry as learning, and inquiry as teaching. Currently, science teachers conceive and practice inquiry in school science as hands-on activities, experiments, or processes of science. These approaches represent progress in science education because they engage students in data-collection strategies; science teachers, however, are less successful in engaging students in the manipulation and analysis of data to develop explanations for the objects, events, and phenomena investigated.

The NSES will present an expanded notion of inquiry in school science programs. Basically, the view of scientific inquiry that will be presented in the NSES places more emphasis on the management of information and ideas than on the management of materials and equipment to develop skills. More than three decades ago, BSCS pioneered the concept of inquiry-oriented curriculum and instruction in biology. Even though inquiry teaching (as described by the NSES curriculum committee) is not evident in most science programs, for the past 35 years BSCS materials consistently have expanded the vision of what inquiry means for subject matter, teaching, and learning. The expanded notion of inquiry, even though it may seem evident and small, will require educators to modify approaches to science teaching.

The NSES recommends that science curricula *include connections with other subject areas*. The National Council of Teachers of Mathematics (NCTM) developed standards (NCTM 1989) that parallel the reform of science education, including using technology, using relevant applications, and having instruction foster active student involvement. Several reports (Bybee et al. 1990; NCTM 1989; Minnesota Mathematics and Science Teaching Project 1973; AAAS 1989) discuss the need to integrate science and mathematics. Other reports (Bybee et al. 1992) recommend integrating science with social studies. When using a problem-centered approach to studying science, other disciplines become an integral part of the study. For instance, the work done at Vanderbilt University on the Jasper series (The Cognition and Technology Group at Vanderbilt 1990) is an excellent example of how science, mathematics, and technology are integrated. To solve the overall problem posed on a Jasper optical disk, students must have information and solve mathematics, science, and technology subproblems.

Since the early 1980s, the science-technology-society (STS) theme has emerged as an important part of the contemporary reform of science education (Bybee et al. 1992; Bybee 1986; Harms and Yager 1981; Hurd 1986; Roy 1985; Rubba 1987). The NSES recommendations express this concern by calling for *connecting science with human affairs*. Such an orientation means the development of curriculum and instruction for the following needs:

- Presenting science knowledge, skills, and understanding in a personal and social context.
- Including knowledge, skills, and understandings relative to technology in the curriculum.
- Extending the inquiry goal to include engineering processes such as cost-risk-benefit analysis and decision making.
- Clarifying the knowledge, skills, and understandings relative to the STS theme that are appropriate to different ages and stages of development.
- Identifying the most effective means of incorporating STS issues into extant science programs.
- Implementing STS programs into school systems.

Instruction. The change toward approaches to instruction reflecting constructivist views about learning is closely linked with the reform of curriculum standards. Up to now, the design of schooling typically reflected a metaphor of an industrial assembly line. The administrators were managers, the teachers were the workers, and the students were the product. You might imagine students rolling down an assembly line with teachers opening up the heads and pouring in the content and skills. In contrast, constructivist views of learning place the emphasis on the student as worker and teacher as manager/facilitator (like a manager in the information industry). The student is the one who does the learning. Constructivists find it unproductive to think of students as *black boxes* for which instructional inputs lead to predictable outcomes (performance on achievement tests). Constructivists are interested in what goes on in the student's mind. The emphasis is placed on helping the student construct meaning fom educational experiences.

Constructivist learning theory suggests that students learn best when they are allowed to construct their understanding of concepts. We base the phrase *constructing their understanding* on a description listed in

- Learning is a natural process that is active, volitional, and internally mediated.
- The learner seeks to create internally consistent, meaningful, and sensible representations of knowledge.
- The learner organizes information in ways that associate and link new information with existing knowledge in memory in uniquely meaningful ways.
- Higher order strategies for thinking about thinking facilitate creative and critical thinking and the development of expertise.
- The depth and breadth of information processed, and what and how much is learned and remembered, is influenced by (a) self-awareness and beliefs about one's learning ability (personal control, competence, and ability); (b) clarity and saliency of personal goals; (c) personal expectations for success or failure; (d) affect, emotion, and general states of mind; and (e) the resulting motivation to learn.
- Individuals are naturally curious and enjoy learning in the absence of intense negative cognitions and emotions.
- Curiosity, creativity, and higher order thinking processes are stimulated by learning tasks of optimal difficulty, relevance, authenticity, challenge, and novelty for each student.
- Learning is facilitated by social interactions and communication with others in a variety of flexible, diverse, and adaptive instructional settings.
- Learning and self-esteem are heightened when individuals are in respectful and caring relationships with others.
- Beliefs and thoughts, resulting from prior learning and based on unique interpretations of external experiences and messages, become each individual's basis for constructing reality of interpreting life experiences.

figure 1 from the American Psychological Association (1992, 1-6).

Use of a constructivist approach ensures that children are active in the learning process. In most textbook programs, students are passive learners. They acquire information by reading about science or by participating in experiences for which the answers are given on the next page of the book. Such learning is meaningless because it does not relate to what students have observed, or experienced, or otherwise already know or have judged to be true.

Meaningful learning does take time. If students are truly to understand the world, they cannot simply read, memorize, and recite isolated bits of information and vocabulary words. They must take time to wrestle with new ideas, to discuss their ideas with their classmates and teacher, to collect data and use that data to draw conclusions,
and finally to relate what they are learning to the world around them.

Science learning is a communal activity. Students learn science through comparing data from investigations of natural phenomena, comparing results and conclusions, negotiating among themselves meaning of personal explanations, and eventually comparing personal explanations with scientific "textbook" explanations. Teachers should establish a science culture in their classrooms where students internalize the values and norms of science, such as withholding judgment, basing conclusions on data, and respecting others' ideas.

Assessment. All too often efforts to improve science teaching exclude one of the driving forces for science programs—assessment. The national reform effort recognizes that assessment is a critical component of science education reform (AAAS 1989; Raizen et al. 1990; Pelavin Associates 1991; Malcom and Kulm 1991; Lawrenz 1991). Most current assessment tools, however, are designed to measure the educational outcomes of the past, not those of the current reform movement. Leaders in education are concerned that current standardized tests used to assess student and program outcomes are inadequate measures of the most important outcomes of an effective science program. Science education reform currently emphasizes the learning of major conceptual themes rather than factual information. Because nearly all current assessment instruments primarily use multiple choice, true-false, and matching questions, these instruments most effectively measure the lower levels of Bloom's taxonomy (knowledge, comprehension, application). Assessment instruments that address the outcomes of higher levels of thinking, understandings of major conceptual themes, and the ability to apply science understandings and approaches to solving real-world problems unfortunately are not very common.

Authentic assessment is the phrase used by those in the forefront of redesigning assessment strategies. According to Frances Lawrenz (1991), authentic assessment involves maximizing the congruence between the desired outcomes of the program and the assessment procedures. Lawrenz suggests that in addition to multiple choice tests, authentic assessment procedures include (1) essay tests, (2) practical assessment, (3) portfolios, (4) observations and interviews, (5) dynamic assessment, and (6) projects.

Parents, taxpayers, and bureaucrats rightfully demand accountability for investments in educational improvement; they want simple, understandable indicators of educational achievement. The current "crisis" in education has been fueled by indicators of poor performance on national and international assessments of educational achievement. Taxpayers and elected officials, therefore, expect educational reforms to directly relate to improved performance on assessments.

The challenge to educational leaders is to produce assessment instruments and procedures compatible with contemporary reforms in curriculum and instruction and that taxpayers will accept as valid indicators of achievement. If we continue to assess the effects of reforms in science and mathematics with instruments and procedures that are designed as valid measures of outdated goals, then we are in danger of promoting public misperception (and lack of support) of the success of the reform effort.

Professional Development of Teachers

As part of the new guiding metaphor of the education system as a *learning community*, teachers are viewed as professionals who engage in continuous decision making about how and when to intervene to facilitate student learning. Previous views of teacher education focused on *training* teachers to perform generic, isolated skills and behaviors (i.e., questioning skills, wait time, direct instruction). Contemporary views of teacher education take a constructivist approach to the *development* of content-specific knowledge and strategies. The focus is on development rather than training, because the belief is that teaching is an activity in which teachers make specific decisions about what action to take in response to a unique learning situation; it is ineffective for teachers to be trained to respond to a limited set of situations, but teachers can develop the knowledge base to analyze a particular learning situation and to chose from a repertoire of strategies to

Figure 1. Guidelines for learner-centered instruction

promote student learning.

Knowledge bases. Teachers regularly make decisions about what and how to teach—as often as one decision every two seconds. In making these decisions teachers draw upon a variety of knowledge bases. Figure 2 lists the most important knowledge bases for teaching.

Constructivist approach. The Biological Sciences Curriculum Study (BSCS) believes that a constructivist approach to learning is appropriate not only for elementary students but for their teachers as well. Teacher development rather than teacher training is the appropriate focus of teacher education. We would like the teachers to become reflective practitioners (Clift, Houston, and Pugach 1990; Cruickshank 1990; Grimmett and Erickson 1988; Mohr and MacLean 1987; Schon 1991) who are empowered to study and implement improvements to their instructional practice (content and pedagogy). Professional development programs might use the strategies listed in figure 3 to promote reflective teaching.

For changes in teaching to occur, teachers must learn about and experiment with the new pedagogy, such as a constructivist approach to learning, cooperative learning, and advanced educational technology (Joyce and Showers 1988; Little 1982). Teachers also need to improve their pedagogical content knowledge, that is, how to interpret science content for students (Shulman 1986). Furthermore, because new approaches to teaching and learning rarely occur without the active leadership of district-level administrators and principals, educational leaders should employ a comprehensive approach to staff development that includes not only the development of teachers but also the development of leaders for change.

The professional development program. The professional development of teachers should be a career-long, seamless program with the foundation established in undergraduate liberal arts courses and subject-matter courses interconnected with education courses, applied and elaborated in extensive field-based classroom work, extended through a multi-year internship with mentoring from master teachers, and sustained throughout the teaching career in continual professional growth, culminating for some in programs to prepare master teachers. Schools and universities are collaborating to achieve this vision through what are called *professional development schools*, where university faculty and school teachers work together to improve teaching, not only of the prospective teachers but also of the teaching staff in the participating schools. The thought is, if first immersed in a school culture where the university faculty and school teachers collaborate on equal footing to study and construct effective educational approaches, prospective teachers will internalize a habit of mind and behavior that will enable

Subject-matter content. The standard for knowledge of subject-matter content traditionally has been that science teachers will complete approximately the same undergraduate courses as science majors. Educational reformers criticize that courses for science majors who are preparing for graduate work in science are not appropriate for teachers who have the task of interpreting science knowledge for students. Beyond the typical science major, science teachers need greater understanding of (1) the history and nature of science and technology, (2) a variety of science and technology disciplines, (3) content specific to the curriculum taught in precollege science and technology, and (4) applications of science and technology to everyday life.

Learning theory. Effective teachers construct their own understanding of how students learn science. They call upon formal theories of learning (i.e., behaviorists and constructivists) and selectively employ instructional techniques based on a personal interpretation of contrasting theories. Effective teachers mediate their interpretation of learning theories with wisdom derived from teaching practice. They understand how students learn and the capabilities and limitations of their students.

Curriculum. Effective teachers have a diverse and deep knowledge of curricula. They have at their fingertips a wide range of effective learning activities from a variety of sources. They can compare and contrast different approaches to curriculum organization (thematic, topical, concepts). They can compare and contrast different philosophies to teaching and learning embodied in different curricula.

Pedagogy. Effective teachers know and can perform a wide range of instructional techniques, including advanced educational technology. They are knowledgeable of and can apply findings from research on teaching (such as questioning skills, wait time, direct teaching, inquiry, and instructional models). They can select the appropriate instructional technique for the particular learning situation (i.e., constructivist approaches to promote conceptual learning).

Pedagogical-content knowledge. Recently, educational researchers have constructed a new term for a critical knowledge base of effective teachers. Shulman (1986) noted that effective teachers apply specific instructional techniques to help students learn particular science content. The expert teacher is aware of typical misconceptions that students might have developed from prior experiences and know activities and explanations that encourage students to improve their understandings.

Figure 2. Knowledge base for teaching.

them to continue their lifelong pursuit of excellence in teaching.

How does one help science teachers develop? First, science teachers need to have a thorough understanding of the nature of science—the activity of science, the culture of science, the process of science, and the product of science. Science teachers also need to learn how to learn science well and to construct an indepth understanding of the science they are to teach, not just a broad overview of topics and a collection of specific facts. Science content courses for teachers, therefore, need radical revision to emphasize what is most worth knowing in science for a science teacher, to model effective approaches to teaching and learning science, and to engage teachers in doing science.

Second, science teachers need to develop understandings of how to facilitate science learning by children and young adults. Science education courses need to provide concrete cases of how students learn science and ways to help students understand specific scientific concepts. Teacher development programs need to help teachers acquire instructional strategies and become familiar with a

Reflection on learning: teachers use interviews of students, concept mapping, reflective note taking, analysis of case studies, and small group discussions to reflect on their own learning and students' learning.

Reflection on self: teachers keep a journal, write a personal biography, and develop a metaphor for their own teaching style.

Reflection on action: teachers conduct case study research in their own classrooms and use microteaching, videotapes of their own lessons, observations of expert teachers, study groups, peer coaching, and mentoring.

Reflection on program improvement: teachers interpret results from interviews of students, parents, and other teachers, innovation configuration checklists, and student outcome data.

Figure 3. Types of reflective practice

diversity of science programs, materials, and learning activities. Science teachers need to see science classrooms where the culture promotes science learning, embodied in the notion of a learning community (Marshall 1990). Finally, science teachers need continued education and mentoring throughout their career to provide new ideas, guidance, encouragement, and support in the pursuit of continuous improvement in their profession.

BSCS, with support from the National Science Foundation (NSF), is applying these ideas for professional development in a large-scale teacher development project—the Colorado Science Teacher Enhancement Program (CO-STEP). In CO-STEP, BSCS is establishing six teacher development centers in Colorado. Each center has the responsibility to provide long-term development and support for science teachers in the upper elementary grades. Each teacher commits 3 years to the professional development program, culminating in the opportunity to receive a master's degree in Elementary Science Education. These resulting master teachers design and implement a change project to help fellow teachers improve the elementary science program in their schools.

One of the most difficult problems facing teacher educators today is how to present the emerging vision of effective science teaching and learning. Teachers are hard pressed to find concrete models of the current vision for effective science teaching and learning. Because the vision is in the process of emerging, only a few science classrooms can be found to use as models. In response to this need, BSCS, with support from NSF, recently started a project to develop video cases of teaching that model the new approaches to science teaching and learning embodied in the emerging vision. The resulting product will be teacher development modules, supported by video on laser disk of science classrooms, focusing on effective approaches to curriculum, instruction, assessment, and equitable teaching.

Factors related to educational change. Educational change is a long and complex process that often begins with the decision to adopt a new curriculum or approach to teaching. The decision to change is only the beginning; Hord and Huling-Austin (1986) found that it takes 3 or more years for teachers to make a substantial change in teaching.

Change requires the personal commitment of the teachers. Consequently, a number of researchers (Beall and Harty 1984; Berman and McLaughlin 1977; Fullan 1982; Rogers 1983; Bandura 1977; Smith 1987; Fullan, Miles, and Anderson 1988; Rogers and Shoemaker 1971; Doyle and Ponder 1977) have studied factors related to a teacher's predisposition for change (figure 4).

In addition to the factors influencing a predisposition to change, researchers (Fullan, Miles, and Anderson 1988; Ellis 1989; Ellis and Kuerbis 1987; Kuerbis and Loucks-Horsley 1989; Edmonds 1979; Kelley 1980; Leithwood and Montogomery 1981; Brickell 1963; Emrick and Peterson 1978; Fullan 1982; Loucks and Zacchei 1983; Meister 1984; Sarason 1971; Becker 1986; Yinn and White 1984; Goor, Mehmed, and Farris 1982; Gray 1984; Grady 1983; White and Rampy 1983; Watt and Watt 1986; Winkler and Stasz 1985) also have identified factors that influence successful change (see figure 5).

Self efficacy. The teacher must have confidence that he or she can successfully implement the new materials and teaching practices.

Efficacy of change. The teacher must believe that the change will improve teaching, ease some teaching tasks, and improve student learning.

Practicality ethic. The teacher must believe that the costs of changing his or her teaching behaviors and materials ultimately will be less than the benefits gained from changing.

School culture. The teacher must perceive that the change is simple to master and implement, that he or she can experiment on a limited basis in a low-risk environment, and that he or she will receive positive feedback from others for changing.

Curriculum fit. The teacher must believe that the change will become part of the established curriculum and that it is not a fad.

Figure 4. Factors related to predisposition to change

Related to Development and Consultation Support

- The teacher must participate in quality training activities.
- The teacher must receive followup consultation, support, and encouragement. The teacher must have the opportunity to practice using the new materials and teaching strategies with individual feedback (coaching) back in the classroom.
- The teacher must provide feedback about the implementation project and about his or her use of the innovation.
- School systems must use that feedback from teachers to plan additional inservice and assistance, to provide supportive materials, and to consider possible modifications in plans, organizational arrangements, and the innovation itself.
- The teacher must have a clear picture of how the innovation can improve science teaching.

Related to School District Support

- The school district must give the teachers time to participate in training, to plan lessons, to review educational materials, and to collaborate with fellow teachers.
- The school district must provide the teachers and students easy access to necessary equipment and materials.
- The central office of the school district must sanction and clarify the need for the innovation, give clear and consistent communication, apply pressure, and provide consultation, release time, materials, and resources for training.
- The school district and building administrators must collaborate with teachers in developing a clear, long-range plan for implementing the innovation in the schools.
- The school district must form building implementation teams that have a shared vision of the change process, agree on and conduct a clear plan for implementation, provide technical coaching and assistance, arrange training, reinforce attempts to change, and put the program in the spotlight for everyone in the school community.
- The school district must provide incentives and psychic rewards to teachers, including special recognition, release time, salary credit, and technical support.
- The school board and community must support the need for innovation.
- The principal must take an active role in initiating, sanctioning, supporting, and responding to the innovation. The principal must provide teachers with access to resources, training, and assistance from others.
- The principal must establish in the school a positive environment conducive to change. The teacher must feel able to explore new approaches and to risk failure.
- The teacher must agree with administrators and other participating teachers on the need, appropriateness, and priority of the innovation.
- The teacher must be involved in designing the implementation plan, selecting the educational materials, designing the instructional units, organizing the equipment and materials, scheduling the use of the materials, and training other teachers.

Figure 5. Factors influencing successful change

BSCS has investigated the factors related to successful change. During the past 8 years, with support from NSF, the BSCS *ENLIST Micros* program (Ellis 1989) has evolved through feedback from field testing the professional development strategies in 18 school districts with more than 300 teachers and through continually updating the program by applying research findings from other studies. Several studies (Wu 1987; Stecher and Solorzano 1987; Smith 1987; BSCS 1989; Stasz and Shavelson 1985) have confirmed the factors listed in figure 6, which are employed in the *ENLIST Micros* program, as characteristics of successful professional development programs.

Teachers need followup in the classroom (coaching) to change their teaching behaviors. Several researchers point out that peer coaching is a cost-effective way to improve teacher training (Leggett and Hoyle 1987; Joyce and Showers 1987; Showers 1985; Munro and Elliott 1987; Brandt 1987; Neubert and Bratton 1987). Garmston (1987) points out that collegial coaching refines teaching practices, deepens collegiality, increases professional dialog, and helps teachers think more deeply about their work. The coaching should be conducted by pairs of teachers; focus on the priority set by the observed teacher; gather data about the teaching strategy, student behaviors and outcomes, and teacher behavior; and help analyze and interpret the data from the observation. It is important that the teachers practice the new strategies in a series of several followup sessions. Showers (1985) and Leggett and Hoyle (1987) recommend these followup activities that fellow teachers might provide on a weekly basis: observing the teacher practice the behavior in the classroom, followed by a postobservation conference; providing support and encouragement; assisting in planning future lessons; organizing sharing sessions for the teachers to discuss successful and unsuccessful lessons; and helping with the location and production of materials.

- Voluntary participation by teachers.
- Multiple training sessions offered over an extended period of time (one semester or more), coupled with followup support in the classroom.
- Credible and knowledgeable instructors. (Teachers often prefer other teachers as instructors.)
- Ongoing involvement of teachers in planning the course. (The instructors must be flexible and willing to adapt the course to the needs of the teachers.)
- Inservice activities matched to the experience and concerns of the teachers.
- Extensive hands-on practice with the materials and teaching strategies that progresses from simple to complex exercises.
- Experience with instructional applications that offer promise for improving science education.
- Comfortable, relaxed, low-risk environment.
- Appropriate balance between lecture and guided practice.

Figure 6. Characteristics of successful programs

Restructuring of the Educational Support System

Effective reform efforts recognize that the whole educational support system must be designed to support the reforms in curriculum and instruction made by the teachers. All educators who are involved in schooling must participate in generating and supporting the reforms; in that way, they become active members of the educational community with the commitment and responsibility for enacting the reforms. Master teachers are effective as educational leaders in individual buildings to encourage and provide technical assistance to other teachers who are implementing the reforms. In addition, principals, district-level administrators, and science education faculty should understand, guide the construction of a shared vision of, and be supportive of the new curriculum, approaches to pedagogy, and effective strategies for fostering change (Fullan, Bennett, and Rolheiser-Bennett, in press).

The educational support system must be responsive to the challenges of educational change. For any innovation to become integral to a school's instructional program, the school personnel must complete the cycle of change: initiation, implementation, and institutionalization. Each phase is critical to the long-term success of any new program initiative because what happens during one phase influences what happens during subsequent phases. More important, successful change efforts include a plan for the activities of all three phases from the outset.

Initiation. Initiation establishes the impetus for change. The events that occur during initiation have a profound effect on the eventual outcomes of the innovation. During the initiation phase, schools establish a leadership team (including external consultants, the principal, a district administrator, master teachers, and parents) to envision, guide, and support change. The leadership team begins by establishing a culture promotive of change where teachers are encouraged to try out new ideas. Over time, a shared vision of the desired change gradually emerges, and the leadership team delineates its philosophy and features. Once the desired change has been identified, a few master teachers might pilot test the innovation. As a result of the pilot test, the district staff (teachers and administrators) collectively would decide whether or not to adopt part or all of the components of the innovation throughout the school system, and the leadership team would design a plan for supporting the implementation of the proposed change.

Marshalling a broad base of support for the innovation is the critical task of the leadership team during initiation. The school improvement program will have a long-term impact on teaching only if district administrators, master teachers, and principals are central to the planning of the implementation of the innovation from the outset (Berman and McLaughlin 1977; Fullan and Stiegelbauer 1991). During this phase, the leadership team asks questions: How can we build a shared vision? How does this proposed change help us achieve our goals? How can we design and establish a comprehensive program for professional development? How can we establish a school culture fostering continual change? What are our long-range plans for change? How can we ensure that the changes become lasting?

Implementation. Implementation, the phase in which teachers begin to use the new approaches to curriculum and instruction, requires at least 3 to 5 years, during which time the leaders for change take many actions to support teachers. If these actions are not part of a strategic plan for supporting change, the innovations probably will not become integral to a school's instructional program. Essential to this plan are activities for professional development, consultation, support, and monitoring of the program's implementation. These activities should be performed by all members of the district implementation team, composed of the principal, a district administrator, master teachers, and the external consultants. The school-based team (principal and master teachers) provides the ongoing and daily support that teachers need to change. For example, the principal ensures that teachers have the materials they need (i.e., supplies, equipment, and software) and consults with teachers about the program, while the master teachers help their colleagues reflect on teaching and learning, plan instruction, and solve problems. The consultants external to the school—the district administrators and university faculty—provide comprehensive professional development emphasizing the latest trends in science education, appropriate uses of educational technology, and strategies for school change.

Institutionalization. The most significant failure of past attempts at educational reform has been the lack of attention to the institutionalization of the changes. The reform is not complete until the changes are no longer seen as innovations, but are accepted as a routine part of schooling. For institutionalization to occur, the members of the leadership team must consider how they will ensure that changes are widespread. Institutionalization requires no less effort on the part of the leadership team than initiation or implementation, but the activities are qualitatively different. During this final phase of implementation, teachers need support to integrate the reforms into other areas. Furthermore, plans for staff development must include strategies to educate new teachers and to enhance the skills of teachers who have begun using the innovation.

Revision of School Culture

School culture perhaps is the most neglected component of reform. Far too often, educational researchers and reform leaders simplify the process of educational change by identifying a formula for successful reform (Fullan and Miles 1992). They list caveats of successful educational change efforts. These caveats are useful and often are derived from research and the wisdom of practice. Adhering to such narrow admonitions, however, focuses attention away from the bigger picture of educational change. No matter how successful and effective the teacher training program, it is unlikely that the reform will be fully implemented or institutionalized if the school culture is not supportive of the specific reform and of change in general.

Successful schools establish a culture fostering educational reform. They engage teachers, parents, administrators, and students in constructing the vision of the reform. They share the decision-making authority among all stakeholders (parents, students, teachers, administrators). They recognize that even though the specifics of the reform may be delineated at the national, state, and local levels, change is done by teachers in their classrooms.

School culture that is supportive of reform recognizes that systemic change is a group process in which individuals together learn new ways of educating. Change is stressful, challenging, and ultimately rewarding. Teachers need to be encouraged and supported in taking risks; trying out a new approach to teaching the first time may lead to failure, but learning new ways to teach can occur only in a culture that accepts failure as a natural part of learning.

It takes a long time (several years) for reform to progress through the stages of initiation, implementation, and institutionalization. Change does not take place when President Bush pronounces that U.S. students will be number one in the world in science and mathematics by the year 2000. It takes place when teachers negotiate the process of change, learning new approaches to education. Change is a continuous process. Successful educational leaders understand that change is a process of building consensus for a common vision of what good teaching and learning look like. It is useful to think of educational change as a journey rather than an engineering task (Fullan and Miles 1992). Engineers use blueprints to establish detailed specifications for the final product, while journeys follow road maps that have multiple paths to the destination. Throughout the change process, the new ideas about teaching and learning grow and evolve within the unique school culture.

Schools must accept that change consumes resources. Change demands a great deal of time from all school personnel; change also requires a large investment of material resources. A nation seeking to reform schools must be prepared to dedicate a large portion of available resources over a period of several years to institutionalize successfully the new approaches to curriculum and instruction.

Conclusion

The conclusion I reach is that to foster reform in science education the Nation must 1) make a coordinated effort at reforming all aspects of the education system and 2) respect and encourage all stakeholders in actively making the changes. NSF has put into place many pieces that together could achieve a coordinated effort of systemic reform—the State Systemic Initiative and the Urban School Initiatives, the National Clearinghouse for Science Education, and the hundreds of teacher enhancement, teacher preparation, curriculum development, and research projects. Should these projects construct a common vision and a coordinated plan of action, these efforts have the promise of making great strides toward putting the rhetoric (*Science for All Americans* and the National Science Education Standards) into practice.

The key to reform is to understand and to respect the roles and responsibilities of all of the stakeholders. Scientists and science educators, "the experts," are fond of producing sweeping policy statements and curriculum programs that capture their vision for what ought to be. Educational change, however, takes place in individual classrooms by individual teachers responding to their unique situation of students, parents, community, and school. In successful reforms, teachers construct their own vision and adapt the ideas provided by the "experts." Perhaps the role of the experts ought to be to collaborate with the teachers in the schools in constructing a shared vision and in making local decisions about curriculum, instruction, and assessment, rather than to proscribe an elegant formula, which if teachers would just follow, would lead to improved science education.

The slogan from environmentalists to think globally, act locally applies equally well to reform in science education. It is vitally important to construct a clear, shared vision of needed reforms in the system of science education in response to changes in the global society and economy. To be responsive to our rapidly changing society, however, we need local educational systems that foster a culture of continuous change.

REFERENCES

American Association for the Advancement of Science (AAAS). (1989). *Project 2061: Science for all Americans*. Washington, DC: AAAS.

American Psychological Association. (1992). *Learner-centered psychological principles: Guidelines for school redesign and reform.* Washington, DC: American Psychological Association.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.

Beall, D., and Harty, H. (1984). The effects of computer use in science and mathematics education upon the computer literacy of preservice elementary science and mathematics classes. *Journal of Computers in Mathematics and Science Teaching,* 3, 34-38.

Becker, H. J. (1986). *Instructional uses of school computers: Reports from the 1985 national survey*. Issue no. 3. Baltimore, MD: Center for Social Organization of Schools, John Hopkins University.

Berman, P., and McLaughlin, M.W. (1977). *A model of educational change: Rand change agent study*. Santa Monica, CA: Rand.

Berryman, S.E. (1983). *Who will do science?* Special Report of the Rockefeller Foundation.

Biological Sciences Curriculum Study (BSCS). (1989). *New designs for elementary school science and health*. Colorado Springs, CO: BSCS.

Brandt, R. (1987). On teachers coaching teachers: A conversation with Bruce Joyce. *Educational Leadership*, 12-17.

Brickell, H.M. (1963). State organization for educational means. In *Media and educational innovation: A symposium on identifying techniques and principles for gaining acceptance of research results on use of newer media in education.* Lincoln, NE: University of Nebraska.

Bransford, J.D., Sherwood, R.D., Hasselbring, T.S., Kinzer, C.K., and Williams, S.M. (1990). Anchored instruction: Why we need it and how technology can help. In D. Nix and R. Spiro (eds.), *Cognition, education, and multimedia: Exploring ideas in high technology*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bybee, R.W., Ellis, J.D., Giese, J.R., Parisi, L.S., and Singleton, L.R. (1992). *Teaching about the history and nature of science and technology: a curriculum framework.* Colorado Springs, CO: BSCS.

Bybee, R.W., Buchwald, C.E., Crissman, S., Heil, D.R., Kuerbis, P.J., Matsumoto, C., and McInerney, J. (1990). *Science and technology education for the middle years: Frameworks for curriculum and instruction.* Washington, DC: National Center for Improving Science Education.

Bybee, R.W. (1986). The Sisyphean question in science education: What should the scientifically and technologically literate person know, value, and do—As a citizen? In R.W. Bybee (ed.), *Science-Technology-Society 1985 NSTA Yearbook.* Washington, DC: National Science Teachers Association.

Carnegie Commission on Science, Technology, and Government. (1991). *In the national interest: The federal government in the reform of K-12 math and science education.* New York: Carnegie Corporation.

Clift, R.T., Houston, W.R., and Pugach, M.C.. (1990). *Encouraging reflective practice in education.* New York: Teachers College Press.

The Cognition and Technology Group at Vanderbilt. (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher*, 2-10.

Cruickshank, D.R. (1990). *Reflective teaching: The preparation of students of teaching*. Reston, VA: Association of Teacher Educators.

Doyle, W., and Ponder, G.A. (1977). The practicality ethic in teacher decision-making. *Interchange,* 8(3), 1-12.

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37, 15-24.

Education Commission of the States (1983). Action for excellence: A comprehensive plan to improve our nation's schools. Denver, CO: Task Force for Economic Growth, Education Commission of the States.

Ellis, J.D. (1989). An evaluation of a project for preparing science teachers to use microcomputers. *Journal of Computers in Mathematics and Science Teaching*, 9(1), 51-64.

Ellis, J.D., and Kuerbis, P.J. (1987). Encourage the literacy of science teachers in the use of microcomputers (ENLIST Micros). Final report to the National Science Foundation. Colorado Springs, CO: BSCS.

Emrick, J., and Peterson, S. (1978). *A synthesis of findings across five recent studies in education dissemination and change.* San Francisco: Far West Laboratory.

Fullan, M. (1982). *The meaning of educational change.* New York: Teachers College Press.

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

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

Fullan, M.G., and Miles, M.B. (1992). Getting reform right: What works and what doesn't. *Phi Delta Kappan,* 73(10), 745-752 .

Fullan, M.G., Bennett, B., and Rolheiser-Bennett. (in press). Linking classrooms and school improvement. *Educational Leadership.*

Fullan, M.G., Miles, M.B., and Anderson, S.E. (1988). *Strategies for implementing microcomputers in schools: The Ontario case.* Ontario, Canada: The Ministry of Education.

Garmston, R. J. (1987). How administrators support peer coaching. *Educational Leadership*, 44(5), 18-26.

Goor, J., Mehmed, A., and Farris, E. (1982). *Student uses of computers in schools*. Washington, DC: National Center for Education Statistics.

Grady, M.T. (1983). Long-range planning for computer use. *Educational Leadership*, 40(8), 16-19.

Gray, P.J. (1984). *Policy analysis: A tool for setting district computer use policy.* Portland, OR: Northwest Regional Education Laboratory.

Grimmett, P.P., and Erickson, G.L. (1988). *Reflection in teacher education.* New York: Teachers College Press.

Hall, G.E., and Hord, S.M. (1987). *Change in schools: Facilitating the process.* Albany, N.Y.: State University of New York Press.

Ham, S., and Sewing, D. (1988). Barriers to environmental education. *The Journal of Environmental Education,* 19(2), 17-24.

Ham, S., Rellergert-Taylor, M., and Krumpe, E. (1988). Reducing barriers to environmental education. *The Journal of Environmental Education*, 19(2), 25-33.

Hardin, G. (1968). Human ecology: Subversive and conservative. *American Zoology*, 25, 469-476.

Harms, N.C., and Yager, R.E. (1981). *What research says to the science teacher.* Washington, DC: National Science Teachers Association.

Hopkins, D. A. (1985). *A teacher's guide to classroom research.* Philadelphia: Open University Press.

Hord, S.M., and Huling-Austin, L. (1986). Effective curriculum implementation: Some promising new insights. *The Elementary School Journal*, 87(1), 97-115.

Huling, L., and Johnson, W. (1983). A strategy for helping teachers integrate research into teaching." *The Teacher Educator,* 19(02), pp. 11-18.

Hurd, P.D. (1986). Perspectives for the reform of science education. *Phi Delta Kappan,* 67(5), 353-358 January.

Joyce, B., and Showers, B. (1988). *Student achievement through staff development*. New York: Longman.

Joyce, B., and Showers, B. (1987). Low-cost arrangements for peer coaching. *Journal of Staff Development*, 8(1), 22-24.

Kelley, E. (1980). *Improving school climate*. Reston, VA: National Association of Secondary School Principals.

Kuerbis, P.J., and Loucks-Horsley, S. (1989). The promise of staff development for technology and education. In J. Ellis (ed.), *1988 AETS Yearbook: Information Technology and Science Education.* ERIC SMEAC at The Ohio State University: Columbus, OH.

Lapointe, A.E., Mead, N.A., and Phillips, G.W. (1988). A world of differences: An international assessment of mathematics and science. Princeton, NJ: Educational Testing Service.

Lawrenz, F. (1991) Authentic assessment. In *NARST Research Matters - to the Science Teacher.* No. 26.

Leggett, D., and Hoyle, S. (1987). Peer coaching: One district's experience in using teachers as staff developers. *Journal of Staff Development*, 8(1), 16-20.

Leithwood, K.A., and Montgomery, D.J. (1981). The role of the elementary school principal in program improvement. *Review of Educational Research*, 52, 309-339.

Little, J.W. (1982). Norms of collegiality and experimentation: Workplace conditions of school success. *American Educational Research Journal,* 19(3), 325-340.

Loucks, S.F., and Zacchei, D.H. (1983). Applying our findings to today's innovations. *Educational Leadership*, 41, 28-31.

Loucks-Horsley, S., Harding, C. K., Arbuckle, M.A., Murray, L.B., Dubea, C., and Williams, M.K. (1987). *Continuing to learn: A guidebook for teacher development*. Andover, MA: The Regional Laboratory for Educational Improvement of the Northeast and Islands. Oxford, OH: National Staff Development Council.

Loucks-Horsley, S., Kapitan, R., Carlson, M.D., Kuerbis, P.J., Clark, R.C., Melle, G.M., Sachse, T.P., and Walton, E. (1990). *Elementary School Science for the 90s*. ASCD.

Malcom, S.M., and Kulm, G. (eds.). (1991). *Science assessment in the service of reform.* Washington, DC: American Association for the Advancement of Science.

Marshall, H.H. (1990). Beyond the workplace metaphor: The classroom as a learning setting. *Theory into Practice,* 29(2), 94-101.

Meister, G. (1984). *Successful integration of microcomputers in an elementary school* (Report No. 84-A13). Stanford University, Institute for Research on Educational Finance and Governance.

Minnesota Mathematics and Science Teaching Project. (1973). *A final report to the National Science Foundation on fifteen years of effort in development of a coordinated mathematics/science curriculum.*

Mohr, M.M., and MacLean, M.S.(1987). *Working together: A guide for teacher-researchers.* Urbana, IL: National Council of Teachers of English.

Mullis, I.V.S., and Jenkins, L.B. (1988). *The science report card: Elements of risk and recovery*. Princeton, NJ: Educational Testing Service.

Munro, P., and Elliott, J. (1987). Instructional growth through peer coaching. *Journal of Staff Development*, 8(1), 25-28.

National Center for Improving Science Education (NCISE). (1991). *The high stakes of high school science*. Washington, DC: NCISE 1991.

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

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

National Research Council (NRC). (1993). *National science education standards: An enhanced sampler.* Washington, DC: National Research Council.

National Science Board Commission of Precollege Education in Mathematics, Science, and Technology. (1983). *Educating Americans for the 21st century*. Washington, DC: National Science Foundation.

Neubert, G., and Bratton, E. (1987). Team coaching: Staff development side by side. *Educational Leadership*, 29-32.

Pelavin Associates, Inc. (1991). Statewide testing programs. Paper presented to the Secretary's [U.S. Department of Education] Commission on Achieving Necessary Skills.

Raizen, S.A., Baron, J.B., Champagne, A.B., Haertel, E., Mullis, I.V.S., and Oaks, J. (1990). *Assessment in science education: The middle years.* Andover, MA: The National Center for Improving Science Education (NCISE).

Rogers, E.M., and Shoemaker, F.F. (1971). *Communication of innovations: A cross-cultural approach.* New York: Free Press.

Rogers, E. M. (1983). *Diffusion of innovations* (3rd ed.). New York: Free Press.

Roy, R. (1985). The science/technology/society connection. *Curriculum Review*, 24 (3).

Rubba, P. (1987). Perspectives on science-technology-society instruction. *School Science and Mathematics,* 87 (3).

Sarason, S. B. (1971). The culture of the school and the problem of change. Boston, MA: Allyn and Bacon, Inc.

Schon, D.A. (1991). *The reflective turn.* New York: Teachers College Press.

Showers, B. (1985). Teachers coaching teachers. *Educational Leadership,* 42(7), 201-206.

Shulman, L.S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher,* 15, 4-14.

Smith, B. (1987). *Factors that influence teacher microcomputer implementation proneness in middle/junior high school science curricula* (Doctoral dissertation, University of Indiana).

Stasz, C., and Shavelson, R.J. (1985). Staff development for instructional uses of microcomputers. *AEDS Journal*, 19(1), 1-19.

Stecher, B.M., and Solorzano, R. (1987). *Characteristics of effective computer in-service programs.* Pasadena, CA: Educational Testing Service.

Task Force on Education for Economic Growth. (1983). *Action for excellence.* Denver, CO: Education Commision of the States.

U.S. Congress, Office of Technology Assessment (OTA). (1988). *Technology and the American economic transition: Choices for the future* (OTA-TET-283). Washington, DC: U.S. Government Printing Office.

U.S. Department of Education. (1991). *America 2000: An education strategy*. Washington, DC: U.S. Government Printing Office.

U.S. Department of Labor. The Secretary's Commission on Achieving Necessary Skills. (1991). *What work requires of schools.* Washington, DC: U.S. Government Printing Office.

Vetter, B.M. (1988). Look who's coming to school: Changing demographics - implications for science education. In *Curriculum and development for the year 2000.* Colorado Springs, CO: BSCS.

Watt, M., and Watt, D. (1986). Making a difference with computers: Successfully integrating computer tools in the school curriculum. *SIG Bulletin*, 21-23.

White, D., and Rampy, L. (1983). *Delphi study on policy issues in the introduction and management of computers in the classroom*. Research report 90. Bloomington, IN: Agency for Instructional Technology.

Winkler, J.D., and Stasz, C. (1985). A survey of incentives for staff development of computer based instruction. Paper presented at annual meeting of the American Educational Research Association, Chicago, IL.

Wu, P.C. (1987). Teachers as staff developers: Research, opinions, and cautions. *The Journal of Staff Development*, 8(1), 4-6.

Yinn, R.K., and White, J.L. (1984). *Microcomputer implementation in schools*. Washington, DC: COSMOS Corp.

Practices that Support Teacher Development: Transforming Conceptions of Professional Learning

Ann Lieberman
Professor of Education
Co-director, National Center for Restructuring Education, Schools, and Teaching
Columbia University

The current school reform effort is seeking to develop and create not only new (or reframed) conceptions of teaching, learning, and schooling, but concommitantly, a wide variety of practices that support teacher learning. These practices cut into some deeply held notions about staff development and inservice education that have long influenced both educators' and the public's views of teachers. Although there is growing sophistication about the process of restructuring schools and the problems of changing school cultures (Murphy and Hallinger 1993; Lieberman 1995; Lieberman and Miller 1992; Fullan 1982; Hargreaves 1994; Little 1993), there is still widespread acceptance that staff learning takes place primarily as a set of workshops, a conference, or a project with a long-term consultant. What everyone appears to want for students—a wide array of learning opportunities that engage them in experiencing, creating, and solving real problems, using their own experiences, and working with others–for some reason is denied to teachers. In the view of traditional staff development, workshops and conferences count, but authentic opportunities to learn from and with colleagues do not. Traditional venues of large group instruction outside the school are taken as almost the only places where adult learning goes on, whereas learning inside of school as an integral part of school life, or as part of a larger network of people struggling with teaching and learning problems, is neither supported nor taken seriously.

The conventional view of staff development as a transferable package of knowledge to be distributed to teachers in bite-sized pieces needs radical transformation and rethinking. It carries not only a limited conception of teacher learning, but one grounded in a set of assumptions about teachers, teaching, and the process of change that does not match current research or practice (Grimmett and Neufield 1994; Little 1993; McLaughlin and Talbert 1993; Wood 1992).

Learning from History: Questioning Assumptions About Teacher Learning

In 1957 the National Society for the Study of Education published the book *In-service Education 56th Year-Book* (Henry 1957). The importance of the book was not only the comprehensiveness of the treatment of the topic, but the challenge it made to the limited assumptions of inservice education that had dominated the early 20th century (e.g., The Teacher Institute, which brought teachers together for lectures, was the primary method for teachers to learn new ideas). The alternative that Henry proposed was that schools and entire staffs should be collaborators in providing inservice education. This view was suggested by the growing knowledge of group dynamics that linked larger ideas of change to whole school problems (Corey 1953; Parker and Golden 1952). Coupled with the increasing status of teachers at that time, the idea that teachers should be coworkers in their own improvement gained credence and some support in educational circles.

The conflicting assumptions–that teachers learn mainly through direct teaching, rather than by being involved in helping to define and shape the problems of practice–carry with them deep-rooted philosophical notions about learning, competence, and trust that are again at the heart of professional development in this era (Cochran-Smith and Lytle 1990; Darling-Hammond 1993; Hargreaves 1994; Lieberman and Miller 1992; McLaughlin and Talbert 1993).

Teachers have been told all too often that other people's understandings of teaching and learning are more important than theirs, and that their knowledge–gained from their daily work with students–is of far less value (Cochran-Smith and Lytle 1990). Outside experts have often viewed teaching as technical, learning as packaged, and teachers as passive recipients of "objective research."

The contemporary reform movement involves such fundamental issues of schooling that conceptions of knowledge building and teacher learning go far beyond the technical tinkering that has often passed for professional development (Little 1993). The process of restructuring schools places demands on the whole organization that make it imperative that individuals redefine *their* work in relation to how the whole school works. Transforming schools into learning organizations, where people work together to solve problems collectively, is more than a question of inserting a new curriculum or a new program; it involves thinking through how the content *and* processes of learning can be redefined in ways that engage students and teachers in the active pursuit of learning goals–a joining of experiential learning and content knowledge. Teaching as telling, which has dominated pedagogy and the consequent organization of schooling and the way teachers see their work, is being called into question as professional learning for teachers increasingly connects to this reconsidered view of schools.

The ways teachers learn may be more like the ways students learn than we have previously understood. Learning and organizational theorists are teaching us that people learn best through active involvement and by thinking about and articulating what they have learned (Resnick 1986; Schon 1991). Processes, practices, and policies that are built on this view of learning are at the heart of a more expanded view of teacher development that encourages teachers to involve themselves as learners in much the same way as they propose their students do. But what does this actually look like in the pedagogical practice of schools? How can we understand the connections between teacher development and school development ?

Learning by Changing: Teacher Development and School Development

This expanded view of professional learning, of necessity, is both personal and professional, individual and collective, inquiry based and technical (Lieberman and Miller, forthcoming). While we have no definitive road maps that lead us directly to how these dualities are negotiated, we do have a growing body of evidence from some schools that have discovered the power and critical importance of professional development when viewed as an integral part of the life of the school. By studying these schools, we can deepen our understanding of how teachers acquire the experience that encourages them to grow and change in the context of school reform (Darling-Hammond, Ancess, and Falk, forthcoming; Lieberman1995; Murphy and Hallinger 1993).

For example, some organizational and pedagogical changes in these schools put new and experienced teachers together to learn from one another; create common periods for planning so that connections can be made across subject areas; use staff expertise to provide leadership for inhouse workshops or meetings (Lieberman, Falk, and Alexander 1994); have self-contained teams where the organizational structure (a team) encourages constant staff learning (Darling-Hammond, Ancess, and Falk, forthcoming); or develop curricular changes that encourage interdisciplinary studies for short periods of time, involving staff in discussion of curriculum and pedagogy created for short time blocks (Ancess, forthcoming).

Numerous curricular, pedagogical, and assessment approaches to student learning also provide powerful professional learning for teachers, involving them in rethinking their role with students while expanding the way students interact with content and the problems of learning. Many instances of professional learning come about as a result of starting with meetings about subject matter content, pedagogical approaches, new means of assessment, or simply learning (Lieberman 1995). What makes the difference for teachers is that the content of the curriculum, the context of each classroom within the school, and the context of the school itself are all considered, with teacher participation central to any changes to be made in the functioning of the school.

From "Direct Teaching" to "Learning in School"

Most of the inservice or staff development activities that teachers are now offered is of a more formal nature; unattached to classroom life, it is often a melange of abstract ideas with little attention paid to ongoing support for continuous learning and changed practices. By contrast, the conception of teacher development that ties together student-centered pedagogy with opportunities for teacher learning supported by favorable and durable organizational conditions is now being tried in many places (Grimmett and Neufeld 1994; Lieberman 1995). By constructing a continuum of the actual practices that encourage teacher growth, we see that such a continuum involves moving from direct teaching–the dominant mode of inservice–to practices that involve learning in school and out of school. The change from teaching to learning is significant since it implies that teacher-development opportunities must become integral to the restructuring of schools. This will, of necessity, involve strategies and mechanisms that are more long range, more concerned with the interactions of groups and individual teachers, and often original and unique to the particular contexts in which they are invented.

This broader approach moves teachers beyond simply hearing about new ideas or frameworks for understanding teaching practice to being involved in the decisions about the substance, process, and organizational supports for learning in school and to finding broader support mechanisms–such as networks or partnerships–that provide opportunities and innovative norms from groups outside the school.

Because direct teaching is currently much of what the public and many districts consider staff development, it is important that teachers, administrators, and policymakers become aware of new and broader conceptions of professional development. At present many districts have 1-7 days of inservice education in the school year where teachers are introduced to new ideas (e.g., new math standards, new forms of assessment). Some districts run workshops on themes or particular subjects, often hiring consultants to handle the implementation of these ideas. While learning about new ideas that affect both the content and the processes of teaching is important, ideas unrelated to the organization and context of one's own classroom have a hard time competing with the daily nature of work–even when teachers are excited about and committed to them.

If reform plans are to be made operational, enabling teachers to really change the way they work, then teachers must have opportunities to talk, think, try, and hone new practices, which means they must be involved in learning about, developing, and using new ideas with their students. This can happen in a number of ways: *building new roles* (e.g., teacher leaders, critical friends, teacher scholars) (Miller and O'Shea 1994); *inventing new relationships* (e.g., peer coaching, doing action research, etc.); *creating new structures* (e.g., problem-solving groups, school site decision-making teams, descriptive reviews); *working on new tasks* (e.g., journal and proposal writing, learning about assessment, creating standards, analyzing or writing case studies of practice, communicating online over particular topics) (Wood and Einbender, forthcoming; Jervis, forthcoming); and, eventually, *creating a culture of inquiry* wherein

"DIRECT" TEACHING	LEARNING IN SCHOOL	LEARNING OUT OF SCHOOL
• Inspirationals	• Teacher Scholars	• Reform Networks
• Awareness Sessions	• Teacher Leaders	• School/University Partnerships
• Basic Knowledge	• Critical Friends	• Subject Matter Networks
• Initial Conversation	• School Quality Review	• Informal Groups
• Charismatic Speakers	• Peer Coaching	• Collaborations
• Conferences	• Action Research	• Teacher Centers
• Courses and Workshops	• Story Telling	• Impact II
• Consultatations	• Sharing Experience	• NEA and AFT Collaborations
	• Teaching Each Other	
	• Problem Solving Groups	
	• Descriptive Reviews	
	• Portfolio Assessment	
	• Experiencing Self as Learner	
	• Proposal Writing	
	• Case Studies Practice	
	• Standard Setting	
	• Journal Writing	
	• Working on Tasks Together	
	• Writing for Journals	
	• On-line Conversations	
	• School Site Management Team	
	• Curriculum Writing	

Figure 1. Teacher Development and Professional Learning : A Continuum

professional learning is expected, sought after, and an ongoing part of teaching and school life (Lieberman 1995; McClure 1991; McLaughlin 1991; Smith et al. 1991).

What characterizes these examples of professional learning is that their life span is not 1 or 2 days, but that they become a part of the expectation for the teacher's role and an integral part of the culture of the school. Learning and development become as varied and engaging for teachers as they are supposed to be for students. Experiencing and helping to produce new knowledge becomes as compelling as consuming already existing knowledge; in fact, one feeds the other. Being involved as a learner and participant provides openings to new knowledge, broadening the agenda for thought and action. (For example, teachers involved in action research, looking at their own practice, often seek affiliation with their colleagues who subsequently may themselves participate in some form of problem-solving activity.) In important ways, such activities link professional learning that is solo and personal to learning that is also collegial and communal. The descriptions that follow illustrate the connection between teacher learning and the mechanisms to support these in-school efforts.

Learning by Observing Children

The Primary Language Record (PLR), a guide for collecting evidence to aid teachers in understanding how students become literate in the primary grades, encourages teachers to observe student habits and choices as they are involved in learning tasks. Using this guide involves teachers in interviewing parents and students concerning students' study habits and interests both at home and school. It provides them with greater breadth of information about their students, helping teachers to become aware of and plan for student differences in learning styles. Most importantly, by observing children closely (with the help of a guide) teachers see that students learn differently, think differently, and engage with their fellow students in a variety of ways. It does not tell teachers what to do, but rather expands their understanding of what is possible. The PLR enables teachers to better use their own professional judgment to build more effective teaching programs by focusing attention on student strengths. Networks of teachers from New York to California support each other in using this tool to integrate child development knowledge with their observations of their students (Falk and Darling-Hammond 1993).

New Pedagogical Approaches to Subject Matter

New and innovative approaches to subject matter teaching are involving teachers in pedagogical as well as curricular changes. These include the writing process approach, which engages teachers in writing, revising, and polishing their own work to experience what it means to learn to write; whole language approaches to integrating language arts, which involve teachers in planning for blocks of time for students to read, write, listen, and speak–teachers and their students integrating ways of thinking about content and how it is learned often revise their class schedules to allow for larger blocks of work time and more opportunities for students to work together and independently; and the Foxfire approach, which encourages teachers to use students' interests and choices to involve them in planning and carrying out their own learning–students gain skills and subject knowledge as they seek information, write, edit, and produce work in a variety of subject areas using projects of their own making. These new pedagogical approaches encourage teachers to be learners as well as teachers, experiencing themselves the struggle for personal and intellectual growth that is an essential part of the learning process and sensitizing them to the nuances of learning and the needs of individuals and groups.

These approaches to student learning do not downgrade the learning of basic knowledge; they use the interests and abilities of students and teachers to invigorate this learning. Instead of simply memorizing lectures or texts, these approaches involve teachers and their students in using learned skills and abilities to identify and pose problems and to seek perspectives and methodologies that help to find answers to these problems. Inevitably this means increasing student content knowledge since solutions to problems depend on such knowledge and the skills and analytical tools developed in the process.

Strategies for Learning Together

The Descriptive Review, a process that focuses on looking carefully at one student at a time, brings teachers together in a group to talk about particular students that individual teachers are finding difficult to reach or teach. In the process of understanding these difficulties, a teacher tells what he or she knows about the child, and the other teachers then introduce strategies that they have found successful in similar situations. In the process, teachers share their knowledge with one another, learn from one another, and, by extension, take responsibility for the growth and development of all children in the school (Carini 1986).

Learning by Integrating Assessment and Curriculum

Through their involvement in new patterns of student assessment, teachers learn by organizing the curriculum in ways that reflect their rethinking of what students should know and be able to do to demonstrate the breadth and depth of their learning. Portfolios, which are exhibitions of a student's knowledge and skills, embrace diverse forms of expression, including science and social science research reports, constructions, multimedia presentations, original works of art and writing, or dramatic presentations (Darling-Hammond et al. 1993). An important example of this process is the work done in the Central Park East Secondary School (CPESS) linking "Habits of Mind" with portfolio assessment (see Darling-Hammond, Ancess, and Falk, forthcoming, for details).[1] Habits of Mind is defined as a set of five principles that involve examining evidence critically, looking at multiple viewpoints, making connections, seeking alternatives, and looking for meaning. These principles serve as a foundation on which to build a pedagogy that teaches students to use their minds well, to enable them to live socially useful and personally satisfying lives. They form the basis for ongoing discussions about breadth versus depth and core versus individualized knowledge, and for developing the kinds of courses and educational experiences to achieve these ideals. The assessment process, which is integral to this work, uses portfolios as a means to involve teachers to "coach" their students: serving as critics and supporters of their work, connecting them to subject areas, guiding them toward completion of graduation requirements, and always helping them to build habits of mind and work that will last beyond graduation.

The organic relationship between portfolios and the principles of Habits of Mind forms the basis for learning for teachers as well as students. The school involves the faculty in continuous work on the definitions and parameters of core subjects; portfolio content and measurements of competent work; what it means to be a coach, advisor, and supporter of students' work; students' responsibilities for creating, revising, and completing academic work; and the kind and quality of social responsibility and interaction they want with students and their families as well as with each other. Although this is a particularly ambitious example, it shows how significant changing the method of assessment can be to teacher learning and development when this becomes an integral part of the daily work of school transformation and is not seen in isolation from the problems and questions that are a part of teachers' daily lives.

[1]New York State has accepted the portfolio process as part of its major reform "The Compact for Learning." CPESS is one of the Partnership Schools - schools have developed assessment and accountability mechanisms that enable them to pursue innovative models of learning.

The Role of Partnerships, Coalitions, and Networks in Teacher Development: Learning Out of School

Although we have been dealing with professional learning for teachers that takes place inside the school, there is growing evidence that important and potentially powerful organizational arrangements exist outside the school as well. These networks, collaboratives, coalitions, and partnerships offer teachers professional opportunities that are different in quality and kind than those that have been available inside the school or in traditional professional development programs (Lieberman 1986a, 1986b; Lieberman and McLaughlin 1992; Little and McLaughlin 1993; McLaughlin and Talbert 1993; Miller and O'Shea 1994; Puget Sound Educational Consortium 1989, 1990). Unlike most professional development strategies, with a "one size fits all" orientation, networks and/or coalitions and partnerships provide opportunities for teachers to commit in small and large ways to topics that they develop or are of intrinsic interest to them or that develop out of their work (Little and McLaughlin 1991). In addition to formal learning they may, by joining informal groups, develop stronger voices to represent their perspectives, learn to exercise leadership among their peers, use their firsthand experience to create new possibilities for students through collaborative work (Jervis, forthcoming; Puget Sound Educational Consortium 1989, 1990), and perhaps most importantly, develop a community of shared understanding that enriches their teaching while providing the intellectual and emotional stimulation necessary for personal and enduring growth and development (Wood and Einbender, forthcoming; Lieberman and McLaughlin 1992, 674). These important opportunities for teacher development more readily exist in environments free from the constraints of the cultures of university- and school-based educators, providing a level of flexibility and collaborative work not usually possible in existing organizations.

The following examples help us to understand the variety of contexts, contents, and collaborative arrangements possible when teachers are learning outside of the schools in which they teach.

The Southern Maine Partnership

Much can be learned from looking at this 9-year-old partnership between the University of Southern Maine and a group of school districts that has established deep roots in both the schools and the university. Initially bringing together teachers at monthly meetings to discuss research and educational practice, the partnership justified its claim on teachers' time by establishing a neutral forum where teachers learned, asked questions, and talked about their teaching practices and problems in a safe and nonjudgmental environment (Miller and O'Shea 1994, 4). (Initially some of the topics were multiple intelligences, grouping practices for students, and new research on cognition and practices in early childhood classrooms.) The impetus for organizing these initial dialogs came from a university professor who believed that both schools and the university should collectively shape the agenda.[2] Eventually, teachers noted that what they believed and valued and what they practiced were not always in synch. As the Partnership grew, it also helped to establish a core of committed teachers, superintendents, and principals who were energized by the discussions, the seriousness of purpose of the participants, and the growing egalitarian arrangements that permeated the group. The substance and spirit were in turn brought back to their home schools by the participants, serving as the catalyst and impetus for staff learning in these schools.

The Partnership has gone through several different phases: moving from discussions, to reflections on the discussions, to serious work by its members in restructuring schools, to making major changes in the teacher education programs in both schools and the

[2] The partnership was started by Paul Heckman and later developed by Lynne Miller.

university through the creation of Professional Development Schools (Miller and Silvernail 1994). Discussions, conversations, consultancies, networking over particular topics, and teacher-led conferences have all developed and changed over time. This progression indicates that a major strength of the Partnership has been its recognition that it must keep changing the kinds of forums it creates to match the growing and deepening needs of its constituents.[3]

The Foxfire Network

Where the Southern Maine Partnership began as a consciously created partnership between schools and a university, the Foxfire network grew out of a teacher's discovery that in order to interest students in learning in his English class, he had to involve them in areas of their own interest and choice. The dramatic story of how this happened has been recorded elsewhere (Wigginton 1985). But what concerns us here is how one teacher's struggle was transformed into a strategy for the creation of teacher networks to provide professional learning beyond the boundaries of one classroom, one school, and one locale.

Beginning as an outreach program, teachers were invited to participate in classes during the summer. Because they themselves were teachers, the original Foxfire group modeled the techniques that teachers might try with their students during the school year—from encouraging students to choose their own topics to research and write about to involving them in identifying their own learning needs–with teachers serving as guide, coach, and counselor. Understanding that meaningful learning needed to be supported over time, they started networks in a few places where the Foxfire course had been given and where there were growing relationships with groups of teachers. These groups, meeting throughout the year, have become a formal part of the Foxfire Teacher Outreach Program, growing from 5 initial networks to 20. These networks now exist across the country, and while some are connected to colleges and universities, they continue to be centers for professional learning created by teachers for teachers (Smith et al. 1991).

The Four Seasons Network: Authentic Assessment

This network was organized by the National Center for the Restructuring of Education, Schools and Teaching (NCREST) to unite teachers from three reform networks: The Coalition of Essential Schools, Foxfire, and Project Zero. The purpose of the new network was to bring teachers together to learn about authentic assessment by learning from experts and each other, and by creating new modes of assessment in their own classrooms and schools. It was created to be a collaboration to support and encourage teacher participation and leadership in the area of assessment. Teachers from 10 states were brought together initially during two summer workshops, while continuity and support was provided year round through the use of an electronic network. This electronic network enabled teachers to share current stories of practice, discuss their struggles around the creation of portfolios and exhibitions of student work, and give each other support and encouragement for taking risks (Wood and Einbender, forthcoming). Since problems of assessment are crucial to all teaching and learning, this network, by involving teachers from previously existing networks, has helped them all to expand the breadth of their reform work.

These are a few examples of networks and partnerships created to deal with complex educational problems that defy simplistic solutions and pat answers.[4] By bringing groups of teachers together–whether in regard to particular subject areas, articulated principles for reforming schools, new pedagogical techniques, or changed teacher education programs in schools and universities–these networks provided them with access to new ideas and a supportive community for the very difficult struggle of translating these ideas into meaningful changes in teaching and learning in each school and each classroom. In the process, teachers have helped to build an agenda that is sensitive to their contexts and concerns, have had opportunities to be leaders as well as learners, and have often committed themselves to goals that are broader and more inclusive than their initial concerns.

[3]The Partnership now offers conversations, consultants, and networks, each offering different entry points for its members depending upon their needs.

[4]There are many examples of these networks, partnerships, and collaboratives. Some well-known networks include the Coalition of Essential Schools, Accelerated Learning, Foxfire, Professional Development School Partnerships, and the Urban Math Collaboratives.

Breaking the Mold: From Inservice to Professional Learning

This paper has been concerned with the limitations of traditional approaches to teacher development and the new learnings that are informing the field. These might be summarized in this way.

- Teacher development has been limited by lack of knowledge of how teachers learn.
- Teachers' definitions of the problems of practice have often been ignored.
- The agenda for reform involves teachers in practices that have not been a part of the accepted view of teachers' professional learning.
- Teaching has been described as a technical set of skills leaving little room for invention and the building of craft knowledge.
- Professional development opportunities have often ignored the critical importance of the context within which teachers work.
- Strategies for change have often not considered the importance of support mechanisms and the necessity of learning over time.
- Time and the necessary mechanisms for inventing as well as consuming new knowledge have often been absent from schools.
- The move from "direct teaching" to facilitating "in-school learning" is connected to longer term strategies aimed at changing not only teaching practice but also the school culture.
- Networks, collaborations, and partner-ships provide teachers with professional learning communities that support changed teaching practices in their own schools and classrooms.

As opportunities increase for professional learning that moves away from the traditional inservice mode toward long-term, continuous learning in the context of school and classroom with the support of colleagues, the idea of professional development is taking on new importance. For if teacher learning takes place within the context of a professional community that is nurtured and developed from both inside and outside of school, the effects may be not only an expanded conception of teacher development, but the accomplishment of significant and lasting school change.

REFERENCES

Ancess, J. (forthcoming). *Learner-centered accountability in an urban high school: A case study of the urban academy*. New York: National Center for Restructuring Education, Schools and Teaching.

Carini, P. (1986). Building from children's strengths. *Journal of Education,* 168 (3), 13 - 24.

Cochran-Smith, M., and Lytle, S. (1990). Research on teaching and teacher research: The issues that divide. *Educational Researcher,* 19 (2), 2-11.

Corey, S. (1953). *Action research to improve school practices.* New York: Teachers College, Columbia University, Bureau of Publications.

Darling-Hammond, L. (1993). Reframing the school reform agenda: Developing capacity for school transformation. *Phi Delta Kappan,* June, 753 - 761.

Darling-Hammond, L., Ancess, J., and Falk, B. (Forthcoming). *Authentic assessment in action: Case studies*. New York: Teachers College Press.

Darling-Hammond, L., Einbender, L., Frelow, F., and Ley-King, J. (1993). *Authentic assessment in practice: A collection of portfolios, performance tasks, exhibitions, and documentation.* New York: National Center for Restructuring Education, Schools and Teaching.

Darling-Hammond, L., and Snyder, J. (1992). Framing accountability: Creating learner-centered schools. In A. Lieberman (ed.), *The changing contexts of teaching*, 11-36. Chicago: University of Chicago Press.

Falk, B., and Darling-Hammond, L. (1993). *The primary language record at P.S. 261: How assessment transforms teaching and learning.* New York: National Center for Restructuring Education, Schools and Teaching.

Fine, M. (ed.). (1994). *Chartering urban school reform: Reflections on public high schools in the midst of change.* New York: Teachers College Press.

Fullan, M. (1982). *The new meaning of educational change*. New York: Teachers College Press.

Grimmett, P., and Neufeld, J. (eds.). (1994). *Teacher development and the struggle for authenticity*. New York: Teachers College Press.

Hargreaves, A. (1994). *Changing teachers, changing time: Teachers work and culture in the postmodern age*. New York: Teachers College Press.

Henry, N. (ed.). (1957). *In-service education 56th year-book, Part I,* 35-66. Chicago: University of Chicago Press.

Jervis, K. (Forthcoming). *Eyes on the child: Three portfolio stories.* New York: National Center for Restructuring Education, Schools and Teaching.

Lieberman, A. (1986a). Collaborative research: Working with, not working on. *Educational Leadership*, February, 28-32.

Lieberman, A. (1986b). Collaborative work. *Educational Leadership,* February, 4-8.

Lieberman, A. (1992). School/university collaboration: A view from the inside. *Phi Delta Kappan*, 74 (2), 147 - 156.

Lieberman, A. (1995). *The work of restructuring schools: Building from the ground up.* New York: Teachers College Press.

Lieberman, A., Falk, B., and Alexander, L. (1994). *A culture in the making: Leadership in learner-centered schools*. New York: National Center for Restructuring Education, Schools and Teaching.

Lieberman, A., and McLauglin, M. (1992). Networks for educational change: Powerful and problematic. *Phi Delta Kappan*, (73) 9, 673 - 677.

Lieberman, A., and Miller, L. (1992). The professional development of teachers. In M. Alkin (ed.). *The encyclopedia of educational research.* 6th ed., Vol. 3., 1045-1053. New York: Macmillan.

Lieberman, A., and Miller, L. (Forthcoming). *Restructuring teachers, their world, and their work:* New York: Teachers College Press.

Little, J.W. (1993). *Teacher's professional development in a climate of educational reform*. New York: National Center for Restructuring Education, Schools and Teaching.

Little, J.W., and McLaughlin, M. (eds.). (1993). *Teachers' work: Individuals, colleagues and contexts.* New York: Teachers College Press.

Little, J.W., and McLaughlin, M. (1991). *Urban math collaboratives: As the teachers tell it.* Stanford: Center for Research on the Context of Secondary Teaching.

McClure, R. (1991). *Individual growth and institutional renewal.* In A. Lieberman and L. Miller (eds.), *Staff development in the 90s*, 221-241. New York: Teachers College Press.

McLaughlin, M. (1991). Enabling professional development: What have we learned? In A. Lieberman and L. Miller (eds.). *Staff development for education in the 90s*, 61-82. New York: Teachers College Press.

McLaughlin, M., and Talbert, J. (1993). *Contexts that matter for teaching and learning.* Palo Alto, CA: Stanford Center for Research of Secondary School Teaching.

Miller, L., and O'Shea, C. (1994). *Partnership: Getting broader, getting deeper*. New York: National Center for Restructuring Education, Schools and Teaching.

Miller, L., and Silvernail, D. (1994). Wells Junior High School: Evolution of a professional development school. In L. Darling-Hammond (ed.). *Professional development schools*, 28-49. New York: Teachers College Press.

Murphy, J., and Hallinger, P. (eds.). (1993). *Restructuring schooling: Learning from ongoing efforts.* Newbury Park: Corwin Press.

Parker, C.J., and Golden, W. P, Jr. (1952). In-service education of elementary and secondary teachers. *Review of Educational Research*, 22, 193-200.

Puget Sound Educational Consortium. (1989). *Teacher leadership: Contributions to improved practice.* Seattle: Author, University of Washington.

Puget Sound Educational Consortium. (1990). *Teacher leadership: New skills -- new opportunities*. Seattle: Author, University of Washington.

Resnick, L. (1986). AERA Presidential Speech. American Educational Research Association Annual Conference.

Schon, D. (ed.). (1991). *The reflective turn: Case studies in and on educational practice.* New York: Teachers College Press.

Smith, H., Wigginton, E., Hocking, K., and Jones, R. (1991). Foxfire teacher networks. In A. Lieberman and L. Miller. (eds.). *Staff development in the 90s*, 193-220. New York: Teachers College Press.

Wigginton, E. (1985). *Sometimes a shining moment: The Foxfire experience*. Garden City: Anchor Press/Doubleday.

Wood, D. (1992). Teaching narratives: A source for faculty development and evaluation. *Harvard Educational Review,* 62 (4), 535-550.

Wood, D., and Einbender, L. (Forthcoming). *An authentic journey: Emergent understandings about authentic assessment and authentic practice.* New York: National Center for Restructuring Education, Schools and Teaching.

Working, Thinking, and Assessment*

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

Education for all. High standards of performance for everyone. Those words come to us easily and frequently. They have for decades. Yet today, as we enter the last decade of the 20th century, they have a new urgency and a new meaning. Education for all must, from now on, mean a *thinking* education for everyone.

Although the words are familiar, a thinking education for everyone represents a substantial new challenge for those whose task it is to lead education into the next century. For, despite our rhetoric of democracy and equality, we have inherited an education system crafted for the 1920s. At that time, when assembly lines were being perfected in our factories and economic growth seemed to require that most people work at repetitive jobs designed by a few efficiency experts, it made sense to teach only a few future managers, engineers, and public leaders to think and to prepare the majority of students for a future of following directions.

This makes sense no longer. If America is to remain productive and competitive into the next century, we will need to adopt new forms of work. A productive future will be one in which workers throughout our factories and service organizations are responsible not only for *doing* their jobs but also for *designing* their workdays, for understanding what they are doing, and for figuring out how to do it. What is more, people will need a capacity for active learning on the job. Jobs—even within a company or a field of work—are likely to change several times in the course of a working life. People will need to work smarter in the jobs they enter and to learn new jobs throughout their lives.

This heightened demand for thinking will require a very different education from what most students have known until now. Although we have talked for decades about education for thinking, in practice it has been reserved for a relative few: the college bound, students from favored families. But today forward-looking employers are calling for the same abilities to think and reason as the colleges are.

Reorienting the education system toward thinking for everyone will require some very new ways of doing things. For decades, we have focused the bulk of educational effort on developing "basic skills." We have become pretty good at it. We know a lot about how to teach basic, routine skills. But we don't know a lot about how to teach everyone to think. Thinking has been reserved for upper level courses, for the "fast" groups in a classroom, or for children in programs for the gifted. The number of students and teachers involved has been limited. As a result, it has been possible to rely on a kind of self-selection process and on lots of personal initiative on the part of students and teachers. We haven't needed a science or a system for teaching thinking. We will need both in the future.

*This paper originally appeared in *Restructuring Learning: 1990 Summer Institute Papers and Recommendations*, published by the Council of Chief State School Officers, Washington, DC, 1993.

What We Now Know About Thinking

About 20 years ago, psychologists began to seriously examine the nature of thinking. Studies of the cycles of how research affects educational practice show that it takes about 20 years before new ideas become understood well enough to affect educational practice. So we are now at the point at which we ought to be able to take advantage of the science of thinking to redesign education practice.

What have psychologists learned that might be helpful to education? Four principles gleaned from psychology help address misconceptions about thinking: (1) thinking isn't really a "higher-order" activity; (2) thinking isn't really a "skill"; (3) thinking can't be divided into convenient components for teaching and testing; and (4) thinking depends on context.

Thinking Isn't Really a "Higher-Order" Activity

Recall the influential *Taxonomy of Educational Objectives*, which puts thinking and problem solving at the top of a hierarchy of skills. It turns out that this hierarchy is misleading. Research tells us today that even the simplest, most basic learning in any subject requires thinking.

For example, it isn't possible to memorize a list—say, of state capitals—without thinking. Ann Brown, whose article appears in this book [see citation, page 1 of this paper], is a pioneer of this kind of research. She has shown that even a simple memorizing task requires thinking out a strategy and applying it. Perhaps the learner will form some mnemonic, such as making up a story about what makes pairs of items go together. Or perhaps the learner will cluster items that are alike in some way. Whatever the particular strategy, it works best when individuals choose it themselves and structure it to fit the specific occasion.

Even the most rudimentary reading involves thinking. To recognize words, for example, a reader has to sort out an incredibly complex system in which letters stand for sounds. Furthermore, reading even the simplest of messages requires inferential skills. No passages, not even the simplest ones in primers, say everything a reader needs to know to make sense of what is said. Authors depend on readers to understand context and fill in missing links. Inference is everywhere. There can be no reading without it.

What this means for education practice is that we will have to encourage and support thinking even in kindergarten. We can't reserve it for the higher grades or the most talented students, because thinking is a part of learning everything.

Thinking Isn't Really a "Skill"

Cognitive research also shows that thinking is a way of attacking problems, a way of going beyond the information as it is presented and turning it into something that can be learned and used. This kind of mental activity is not a skill in the usual sense of the word. It is not a set of processes that can be practiced on its own. Thinking can't be disengaged from knowledge. In other words, we can't take thinking out of thinking *about* something.

One important implication of the fact that thinking is always embedded in knowledge is that we can't have a separate thinking curriculum, a "process" curriculum as opposed to a "knowledge" curriculum. Learning to think can't be tacked on to an otherwise business-as-usual program of instruction. Experts in the field disagree about whether it pays to develop separate programs for teaching thinking skills. David Perkins, who also writes in this book [see page 1 citation], believes that such courses may be worthwhile. I am more of a skeptic. But even people who disagree on this point still agree that any separate course that is offered will be ineffectual unless it is connected to the rest of the curriculum. Thinking has to be applied to whatever there is to be learned.

That thinking is not a separable skill also means that thinking has to permeate the assessment system. Ideally, every test should be a thinking test. We can't have basic skills tests here and thinking tests there and get what we really want: young people who can think their way through whatever they have to learn, whether they are in school or out in "real life."

Thinking Can't Be Divided into Convenient Components for Teaching and Testing

As educators, we're used to thinking of knowledge and skills as collections of "items." In teacher education and graduate education courses, we learned that knowledge and skill are decomposable into separate components. The way to help students master a complex skill, we were taught, is to teach the separate components. Later, students will put these components together in complex performances.

That notion comes right out of associationism and behaviorism, bodies of psychological theory that predated recent research on thinking. In those theories, the mind is put together in much the same way that a machine is assembled. To build a machine, you first manufacture the individual parts. Then you put them together into subassemblies. Finally, you put the subassemblies together. If everything was done right, the machine runs.

That may be a good theory for machines, but not for minds and not for thoughts. A thought can't be built up out of components. If you analyze thinking into tiny bits and then teach those bits, you get many little bits that never get put together and that probably *can't* be put together. Students who are trained in all kinds of separate thinking skills don't usually think better as a result.

Thinking Depends on Context

For decades, we believed that there was school, and then there was the "real world." School was perceived as a place for getting ready for the real world. In school, students learned an abstract set of general skills and then went forth into the real world and applied them. So you could learn to think in some abstract sense in school and then apply thinking skills outside wherever you happened to be.

Cognitive research tells us that things don't work quite that way. Thinking appears to be highly attuned to the context in which it is done. Abilities learned in one place can't be lifted out of context and used somewhere else. Instead, it looks as if people have to practice in situations very close to the ones in which they'll be using their new competencies. It's much the same for thinking abilities as for athletics. Coaches don't train football players by having them spend a lot of time practicing tennis. Players learn football by practicing football.

The context-dependence of thinking means that educators urgently need to understand the various contexts for which they are aiming to prepare students. Then, we need to make schoolwork a lot more like the "real" work that students are preparing for. That will mean, among other things, more of the tools of tomorrow, more teamwork, more complex tasks for which the teacher doesn't know the single right answer in advance.

What's Wrong with Testing: Decomposing and Decontextualizing

If thinking can't be decomposed into small bits and can't be lifted out of context, then neither can testing. If you want to know how good students are at team problem solving, you have to give them complex problems to solve in teams and observe how well they do. Testing the components won't work. For example, if we want to find out how well students can conduct a debate, it won't do to give them a pencil-and-paper test on detecting errors in arguments. Error detection is only a small part of debating, and working alone on paper produces very different forms of reasoning from arguing verbally with an opponent. The only way to assess debating skills is by observing and judging debates.

The tests we now use, the ones that dominate so much of our school life, are almost the antithesis of the kind of assessment we will need for the thinking programs of tomorrow. The testing technology we have inherited is fundamentally a product of the 1920s and 1930s. Updated in each decade and made more and more elegant technically, American testing has never replaced the basic assumptions of the assembly-line model of knowledge.

The model assumed that knowledge and thought could be decomposed into bits and that speed of responding was always a vital measure of competence. So tests were made up of many short, unrelated, rapid-fire questions. To do well on a standardized reading test, for instance, students must answer questions at the rate of about one per minute. In mathematics and some language skills tests, the rate is two to three questions per minute. What is more, it was assumed that short passages and quickly solved problems could stand in for long reading selections and complex interdependent problem solutions. For example, in the three standardized reading tests most widely used in state assessments, the longest passage for 8th- to 11th-grade students is about 600 words—the equivalent of two typed pages. This means that, according to the tests, we can know how well people can read a book by testing how they respond to two typed pages.

Most current tests also decontextualize to an unacceptable degree. Until very recently, if we wanted to know how well students could write, we gave them what were, at best, editing tests. The tests said, basically, "Here is a passage with errors. Find them and fix them." It is true that good writers edit their own work. But editing what you have written yourself is a very different matter from finding errors in a printed passage written by someone else. What you notice, what you care about, how you choose to "fix" errors, all depend on whom you are writing for and what you are trying to communicate. A decontextualized editing test is not a valid substitute for writing.

If you want to know whether students are capable of writing essays, that is what a test should ask them to do. We are, finally, beginning to build assessments to do that. Many states and school districts are introducing writing assessments that call for demonstrated student writing. This is an obvious and sensible approach. It assesses writing competence in an authentic context. There is every reason to hope that the new forms of writing assessment are just the beginning of a new assessment movement that will soon bring testing into accord with today's understanding of the nature of thinking and learning.

Valuing Effort, Not Aptitude

There is another serious problem with today's tests. They undermine effort and teach us to value inherited aptitude over hard work. Thinking is hard. It demands great effort. It takes hard work both to become good at thinking and to use one's thinking capacity whenever it's needed. But our system of testing does not encourage effort. Instead, it fosters the belief that capacities to learn and to think are inborn—or a least learned very young, well before school begins—and that school can do little change them.

Our testing practices are designed under the assumption that tests shouldn't be "taught to." We believe that it is virtually a form of cheating to prepare students directly for tests. We worry that tests will "drive the curriculum." We keep the content of tests as secret as possible. The result is that *American tests do not encourage extended effort toward publicly known and understood goals*.

In this state of affairs, we are almost alone in the world. We are virtually the only country without an examination system as the capstone of its education program. Throughout most of the world, examinations set a known target that students are expected to meet. Schools set up courses of study that enable students to meet the target. Students and their teachers know the broad outline of the examination, although they don't know exactly what questions they will be asked until the exam period arrives.

In an examination system, teachers prepare students to take the exam in much the same way that coaches prepare a basketball team for the test of the game. In other words, teachers become allies of students in a competition that will be graded by someone else. Teachers and students are on the same team; they are not adversaries.

Our system is different. Our tests are supposed to be general indicators of learning, not coupled to any particular course of study. Tests of achievement in math, for example, are supposed to sample math knowledge in general, not any particular curriculum in math. The Scholastic Aptitude Test (SAT) explicitly disavows a connection to the curriculum. High schools do not explicitly prepare students for the SAT, although students whose families can afford it usually take private cram courses that tend to raise scores a certain number of points. Students taking the ACT tests for college entrance are urged to study the school curriculum in preparation. But the recommendation does not point toward any particular curriculum. ACT tests science in general, history in general, English in general.

How did such a system come about? Was it sheer perversity that led us to adopt a testing system designed to discourage effort? No. Our testing system is rooted in a theory of fairness and equity that was itself based on psychological theories of the 1920s. The reasoning was that we wanted to enable students who didn't have access to good exam preparation—students then primarily in rural schools—to do well on the test all the same. The way to do that, it was thought, was to tie the test to no specific courses of preparation. Then students with high ability to learn could be picked out and given chances for excellent further education.

The idea made sense if one believed that talent was basically inherited, that ability to learn was more a matter of genes than of hard work. That is what psychologists and most educators believed in the 1920s. But we now know that ability can be created by effort and that no test can measure pure talent or pure aptitude. Doing well on a test, even one called an aptitude test, is highly dependent on what the test taker has learned. Some people learn more easily than others. But, if they haven't had a chance to learn and if they have not exerted the effort required to learn, they will not do well, even on the SAT, not even on an IQ test.

Once we admit this, the rationale for decoupling tests from curriculum disappears. Fairness cannot lie in concocting tests that are supposed to let talent shine through without preparation. Fairness and equity can lie only in providing the preparation needed to do well on tests that we know depend on learning. Fairness, in other words, requires a system that recognizes and rewards persistent hard work.

Besides an antiquated theory of fairness, there is another reason that we do not deliberately prepare American students for examinations. We are committed to local control of education. In theory, that means that every school district has its own curriculum. By not linking tests to curriculum, we are able to use only a few tests, thus theoretically allowing curriculum to vary freely.

That is the theory. But, as we all know, that is not how things work out in practice. We have a deep need to compare ourselves with one another—to compare schools, to compare districts, to compare states. For that, we need a common standard. And our small number of achievement tests, all very similar to one another and with carefully worked out statistical comparison tools, seems to provide that standard.

That might not be too bad if the comparisons didn't really matter, if comparative test scores didn't count. But educators know that they do matter—to students, to taxpayers, even to real estate agents. So naturally they want their students to score well, and they do what they can to bring that about.

The result is that we can't enforce our theoretical aversion to teaching to the test. SAT preparation courses are known to raise SAT scores 50 to 100 points, providing unfair advantage to the mostly well-off students who can attend private cram schools and producing calls for SAT preparation within high schools. But is that what we want to spend precious educational time on? The SAT was not designed to organize teaching and learning. Practice on taking SAT items is unlikely to produce the thinking students we are hoping for.

The problem isn't limited to the high school. A recent story in the *New York Times* reveals the dilemma facing caring educators. The article compared two schools in Queens. The two had similar populations of students, but one had reading test scores at the 36th percentile, whereas the other stood at the 80th percentile. Math scores were practically identical, so this couldn't have been a case of an unexpectedly brilliant population in one of the schools. What, then, did account for the difference? It turns out that, in the school with the very high reading scores, the students were practicing for a minimum of one period a day on the exact item forms that were going to show up on the test. They were being prepared for the test, a test that few believed demonstrated real reading competence.

Even the principal didn't believe much in the test. She was the kind of principal whom the "effective schools" research literature teaches us to celebrate, the kind who cares about the students and demands committed teaching from the staff. She was distressed about what she felt she had to do on behalf of her children. She was quoted as saying:

> *I would prefer if we could find a different way of evaluating children's progress, but given the reality, I have no choice. Schools are looked at in this way; children are looked at in that way. I'm not certain that the instrument is testing what we want to test, but what can I do?*

So says a dedicated educator, one who is obviously good at her job, but who is trapped by tests that sprang from a theory of knowledge and of inborn talent that we now know doesn't work.

Our tests were designed on a theory that measurement should not disturb the system. These were supposed to be like thermometers, registering temperature but not changing it. That works for temperature, because the molecules in the air don't care what the temperature is. They don't adjust their movement to a particular temperature reading on the measuring instrument. Educators do care; they adjust their practice to produce temperature readings—test scores—that people value. Education is, de facto, measurement driven. And we wouldn't, on reflection, really want it any other way. We wouldn't want teachers not to care about their students doing well. But we are trapped, because we are using measuring devices that don't adequately represent the knowledge and capabilities we want to teach and that favor aptitude over effort in our thinking about the possibilities for children.

A Different Kind of Measurement-Driven Instruction

Measurement-driven education—which seems to be inevitable in practice, whatever our theories—can work for us instead of against us. It can do this if we use a radically different kind of measurement, a form of measurement designed to be taught to. If we designed our assessments so that they represented well the knowledge and capabilities we value, and if we made the assessments public so all students knew what they should be working toward, we would have the elements of an American examination system. We could begin to make tests work for us instead of against us.

A public examination system is the only way to make American education really fair. Publicly known criteria for success in school would make it possible for all students, and their parents and advocates, to understand what was expected of them. Today, students from inner city and other disadvantaged schools are de facto taught a different curriculum from that taught to children from our more favored schools. The poor are drilled to the 1920s-style tests. The richer schools can afford to ignore them. Indeed, parents and community leaders understand that more is needed and require a more demanding education, something closer to the thinking curriculum, for their children. A public examination system would help to level the playing field, would provide more favorable chances for students who don't come from families in which they can learn what is expected by osmosis. For such less fortunate students, an examination system would make explicit what they were expected to learn. For perhaps the first time, they would know how to prepare themselves.

We cannot build a public examination system on the kinds of tests that we have now. Today's tests would restrict what was studied to bits of information and shallow reasoning. We would get raised test scores if all teachers directly taught students to take the tests, but we wouldn't get better education.

A public examination system requires assessments that directly examine complex performances and that hold decomposition and decontextualization to a minimum. We know three broad classes of ways to do that: performance, portfolios, and projects.

Performance Assessments

We needn't give tests made up of many separate multiple-choice items in order to get the reliable measurements that state legislatures and local boards of education want. It is just as feasible to use ratings of global performances, thus avoiding decontextualization and decomposition.

We know it can be done, even in high-stakes assessment. We do it, for example, in the Olympics. A figure skater in the Olympics does not take a test in skating a straight line, in skating a curved line, or in doing all the separate turns and maneuvers that are the components of figure skating. Instead, the skater performs complex, choreographed routines in which all these elements are present, permitting raters to observe and score them, but always in the context of an integrated performance. To the untrained layman, it is astonishing how close the agreement usually is among a group of Olympic raters. The stakes are high. Countries care about the scores. Careful, professional training makes reliable scoring possible.

The same kind of thing happens in international music competitions. Violinists don't take decomposed tests of bowing, pizzicato playing, and the like. They play Mozart or Bach, and trained raters judge their playing on a number of dimensions. Again, agreement is usually astonishingly high.

We can do the same for school subject matters. A dozen or more states already have writing assessments, for instance, and several states are mounting performance assessments in other subject matters. These efforts show that we can obtain the kind of numbers our political process seems to demand from contextualized and integrated performances.

Portfolios

Even performance exams, however, can't adequately assess all of the competencies called for by the thinking curriculum. Some kinds of student work do not lend themselves to performance on demand. One alternative is a portfolio of student work, building on the model of assessment used in the visual arts. Artists compile their work over time. A painter, for example, may paint for two years and then select the best of the paintings done in that period to submit to a jury for inclusion in a show.

There are three key elements in a portfolio system of assessment: the accumulated work (the paintings), the selection of work to be submitted in the portfolio, and the judging. All three are vital. Obviously, without a collection of paintings, there can be no selection or judging. But selection is also key. A portfolio is not just a collection of any paintings or of all paintings; it is a carefully selected set. The self-selection process is a vital part of the assessment and of the education process. Portfolio assessment thus carries an important implicit message: that assessment, teaching, and learning are all closely linked.

The external judgment process is crucial, too, for it allows teachers to help students select pieces to include; and, in the process, it helps teach students the criteria that they should apply to their own work. For some of the abilities central to the thinking curriculum, the most effective means of examination will be portfolios of work collected over time, selected by students with help from their teachers, and judged by an external panel of teachers.

Projects

Some kinds of learning can occur only by engaging in extended activities with many interdependent components. Adequately supervised and well documented, such projects can serve as both learning opportunities and assessments of learning. A very good example of how a project can serve us both is the Scout system of merit badges. Scouts must earn a specific number and type of badges to qualify for certain ranks. And most of the badges engage Scouts in extended forms of learning that could not be managed in the confines of the regular classroom.

At a recent Educational Testing Service conference, American Federation of Teachers president Al Shanker talked about his memory of the bird identification merit badge he earned as a boy. He noted that, as a real city kid, he didn't like to get up early and didn't much like to walk in the woods. But he was a Boy Scout, and that merit badge was next on his list, so it got him out in the woods at six in the morning. This early morning communing with nature got to be a habit, Shanker said, one in which he sometimes continues to indulge. Thus an "examination" turned out to be a deep and lasting educative experience.

There are forms of project evaluation other than merit badges. Some are particularly useful for developing and assessing skills of working with others. Such skills are, of course, very difficult to measure on just a single occasion, as in an on-demand performance assessment. A more reliable measure might be based on participation in such activities as getting out an issue of the school newspaper. Specific questions can be asked of supervisors and peers that reveal what the student being evaluated was like to work with over time. Thus, a project evaluation doubles as a learning experience, a situation very much like what takes place in evaluating an adult's on-the-job performance.

Speaking of jobs, many high school students hold down after-school, weekend, and summer jobs. Why not use job performance as part of our assessment system? Students' job supervisors know a lot about them. And as apprenticeship systems are added to facilitate school-to-work transition, they too can become opportunities for project-type examinations.

Yet another model of project-based assessment is the science fair. We now use them for a few students; but why not science fairs for all? We would need a somewhat different kind of science fair, one that did not count on parents as available mentors for students as they prepared. But properly organized science fairs could become powerful learning and examining experiences for all students.

Where To Start

How can we establish a true examination system for American schools, with exams that can be studied for and that are graded externally so that students and teachers can work together toward every student's achievement?

The first step is to ask whether we have any elements of such a system already in place. Fortunately, the answer is *yes*. The Advance Placement system run by the College Board is, in fact, a public examination system, although we use it for only a tiny proportion of students. It might be possible to expand the Advanced Placement system rather than start entirely from scratch to build an American exam system. At one time, the New York State Regents operated as a real examination system and may provide another useful model for a nationwide system.

As I have already noted, performance assessments are under development or in use in some states. These may constitute the beginnings of a new approach to assessment in this country. We must be cautious, however. These beginnings must not be taken for the end product. Although they represent a large step forward in the way tests are constructed and used, most state assessments are not examinations in the true sense, because they are not attached to curriculum and students cannot directly study for them.

The National Assessment of Education Progress (NAEP) assessments are not exams, either. They could perhaps be turned into true exams. But it is not clear that it would be a good idea to disassemble what is working well as an indicator system; for, in addition to an examination system, we probably will also need state and national temperature-taking systems for many years to come.

Who Will Build the New System?

Although we have some promising beginnings, we will need substantial new development to get from the current American testing system to the kind of examination system I have outlined here. Who should build this new system? The answer is, whoever agrees that this country needs thinking-oriented and effort-based education. If you, as chief state school officers, agree, then you are the right ones to do it. Statewide exams with cross-state "moderation" would be a natural possibility for America. Individual school districts or consortia of districts could also conduct exams, and "moderation exercises" could be held within states.

Moderation is a technical word drawn from the British public examination system. Until recently, control over education in Great Britain has not been centralized. Specific exams and, therefore, the matching specific curricula are not nationally set. Rather, many different examination boards set and grade exams, and schools choose which exams to prepare their students for. The different exam boards meet for moderation exercises to keep their standards comparable. Teachers from different examining boards sit together to regrade each other's exams and to discuss where they agree and disagree. Over time, this moderation process keeps standards national, even though control of education remains local.

Because instituting an examination system is such a large challenge, it is likely that whoever takes up the challenge will need considerable help. You can get as much help from the research community as you ask for. A national group is now organizing to help design examination systems and, if necessary, help run them. Foundations are also ready to help support this work. We can, therefore, mount the technical assistance needed. Exams need not be a decade away. If a group decided now to begin to build a true examination system, we have enough worldwide experience and enough U.S. technical skill to make rapid progress.

I want to emphasize that measuring and accounting represent a secondary goal of the kind of examination system I am advocating. The primary goal is to convert our aptitude-oriented education system to a system that is effort-based and equitable. The right kind of examination system will help to convince students and teachers that success depends mainly on effort, that early abilities don't control outcomes, and that abilities can be developed through effort and work.

Exams alone cannot produce this change, however. If all we do is put some new exams in place, we will reap failure and social divisiveness. We cannot impose exams without offering education. We can give our state legislators and school boards everything they want in the way of measurement. But the real purpose is better teaching and learning. The whole idea is to *prepare* students for exams. Students have a right to achieve; as educators, we have the obligation to help them to that. Building the assessment system of the future is not a matter for psychometricians and testing experts, but for educators striving to uphold that obligation.

Equity: Providing Equal Access to Powerful Ideas

Richard Lesh, Principal Scientist
Educational Testing Service

Introduction

It is said that more mathematics and science has been created during the past 20 years than in all of the rest of history combined. It is said that to live and work in a technology-based society, foundations for success are not restricted to *geometry* from the time of Euclid, *algebra* from the time of Descartes, *shopkeeper arithmetic* from the industrial age, elementary *logical reasoning* from the time of Aristotle, and *a few isolated topics in science* from the time of Newton. It is said that what is needed in schools is not just new ways to teach old ideas; new levels and types of under-standing also are needed. But the following kinds of evidence suggest that most people, including leaders from business and government, simply do not believe it!

- National *Curriculum and Assessment Standards for School Mathematics* have been defined almost exclusively by school people, with little input from parents or community leaders or leaders from business and industry. Conse-quently, such people tend to be very skeptical about these new standards (Is this another round of the new math?); and most of them expect textbooks, teaching, and tests for their children to be only marginally different the kind that characterized their own days in school.

- It is known that existing national literacy tests are poorly aligned with new national curriculum Standards in mathematics and the sciences. Most focus on little more than low-level facts and skills, familiarity with a few ritualized problem-solving settings, and general creativity and test-taking skills. Yet, such tests are used to measure educational progress; and most people continue to believe that such knowledge and abilities are prerequisites that must be "mastered" before deeper and more powerful understandings and abilities can be introduced.

- People who analyze job requirements are seldom expected to be specialists in mathematics or science. Consequently, lack of expertise tends to prohibit the recognition of mathematical or scientific thinking that goes beyond a narrow and shallow band of rule-following abilities.

Every decade or so in the history of mathematics and science education R&D, the pendulum of curriculum reform swings from emphasizing *behavioral objectives* (BOs: strings of factual or procedural rules) to emphasizing *process objectives* (POs: content and content-independent processes and strategies), or vice versa. Periodically, some attention also is given to *affective objectives* (AOs: feelings, values). Yet in all of the preceding cases, *cognitive objectives* (COs: models for describing patterns and regularities in the world) tend to be almost completely neglected.

The following section of this article describes several recent trends that have led to radical changes in the type of mathematical understand-ings and abilities needed for success in a technology-based society. Based on these trends, I will argue that past conceptions of mathematics (and mathematical learning and problem solving) are far too narrow, shallow, and restricted to be used as a basis for identifying students whose mathematical abilities should be recognized and encouraged. Furthermore, it is misleading to speak of a given test, or textbook, or teaching program as treating students fairly if the situation as a whole reflects biased, obsolete, and instructionally counterproductive conceptions about the nature of mathematics, problem solving, teaching, and learning.

One of the foremost goals of mathematics and science instruction should be to provide democratic access to powerful ideas (cognitive objectives: conceptual models, structural metaphors). Yet, the layer-cake curriculum that characterizes American education sorts topics into artificial categories that destroy both their practical usefulness and theoretical coherence; this curriculum also tracks students into strands that put most powerful ideas out of reach–beyond a facade in which useless and meaningless technical minutia are treated as if they were prerequisites to powerful deeper and more powerful ideas.

How can a broader range of mathematically capable students be identified and encouraged? Our research suggests that the key is to focus on the kind of models and modeling processes that are needed when elementary mathematical systems are used effectively in everyday problem-solving situations. But our research also suggests that new students are not likely to emerge if modeling processes are treated as little more than specific instances of George Polya-style heuristics, strategies, and processes (POs), or if the applications are used mainly as devices to increase motivation and interest (AOs). To see why this is so, it is useful to examine more closely several of the previously mentioned trends that are influencing modern views about the kind of knowledge and abilities that are needed for success in our modern world.

Trends Related to Technology

During the past decade, a revolution has occurred in the core curriculum areas—reading, writing, and mathematics. Technology-based tools are now used on a daily basis in fields ranging from the sciences to the arts and the humanities, in professions ranging from agriculture to business and engineering, and in employment positions ranging from entry level to the highest levels of leadership. Furthermore, these tools have radically expanded (i) the kinds of knowledge and abilities considered to be basic for success in a technology-based society, and (ii) the kinds of problem-solving/decision-making situations that need to be emphasized in instruction and assessment.

In a technology-based society, tools are not simply new ways to carry out old procedures, and they are not simply conceptual crutches to avoid thinking; they are both conceptual and procedural amplifiers that introduce new ways to build (and make sense of) new types of structurally interesting systems, and they also emphasize new types of knowledge and abilities—and new levels and types of understanding for old ideas and abilities. For example, when a business manager uses a graphing calculator or a graphics-linked spreadsheet to make predictions about maximizing cost-benefit trends, these tools not only amplify the manager's conceptual and procedural capabilities when dealing with old decision-making issues, they also enable the manager to create completely new types of business systems that did not exist before the tools were available, and they emphasize completely new types of decision-making issues that need to be addressed. Therefore, in fields ranging from business to engineering, the instruction that is offered by leading professional schools tends to focus heavily on case studies (or simulations of "real life" problem-solving situations) in which technology-based tools are used to create models (or conceptual systems) for generating useful descriptions, explanations, manipulations, and predictions to serve as prototypes (or structural metaphors) for making sense of a wide range of other real life decision-making situations. These models can be thought of as the most important *cognitive objectives* of instruction; and the students who are likely to be the most successful in a given field tend to be those who are able to construct and use the foundation-level models that have the greatest power and the widest range of applicability.

When technology-based tools are available in problem solving, (i) new types of systems emerge as important (such as those involving recursive functions rather than simply isolated computational rules, or sets of data rather than simply pieces of data), (ii) new levels and types of understandings tend to be emphasized (such as those that involve multiple interacting representation systems, spoken language, written symbols, real life metaphors, or a variety of different types of concrete or graphic images), and (iii) different stages of problem solving tend to be emphasized (such as those that involve data selection, organization, and interpretation, or those that involve partitioning complex problems into modular pieces and planning, monitoring, and assessing intermediate results). Consequently, if a broader range of real life problem-solving situations are emphasized, and if one recognizes a broader range of mathematical knowledge and abilities that contribute to success, then a broader range of students will emerge as capable. For this strategy to work, however, it is essential to recognize that all types of mathematical abilities cannot be collapsed into a single trait, such as "general mathematical aptitude" ("g") or some other euphemism for not knowing what is being measured! As long as we continue to collapse all achievements and abilities into a single *score* or *letter grade*, discrimination is inevitable—especially if measurement is based on narrow and distorted beliefs about the nature of "real" mathematics and "real life" problem solving.

Research in mathematics education offers overwhelming evidence that (i) there are many alternative types of mathematical talent, (ii) many different kinds of personalities, knowledge, and capabilities can lead to success, (iii) many different types of success are possible, and (iv) most people have irregular profiles of expertise, with strengths in some areas and weaknesses in others. For example, people with extraordinary spatial/geometric abilities do not necessarily have extraordinary analytic/algebraic abilities (e.g., topologists are not necessarily good tax accountants). Computer programmers (or statisticians) who are very skillful at working within the constraints of one language (or set of paradigms) in traditional kinds of situations are not necessarily skillful at developing new languages (or new paradigms) in nontraditional situations. And, in business and industry, people who are good at working alone to answer other people's clearly formulated questions are not necessarily good at figuring out ways to think about fuzzy situations; they are not necessarily good at dividing complex problems into subcomponents that can be addressed by teams in which the efforts of people with diverse talents and expertise must be coordinated; and they are not necessarily good at adapting to new situations involving new tools and resources.

Trends Related to Psychology

During the past decade, *behavioral psychology* has given way to *cognitive psychology*; one of the foundation-level principles of the newer system is that humans interpret their experiences by developing internal models (or structural metaphors) that enable them to use meaningful patterns and relationships for purposes such as (i) basing decisions on a minimum set of cues (because the model embodies an explanation of how the facts are related to one another), (ii) filling in holes, or going beyond, the filtered set of information (because the model gives a "holistic interpretation" of the entire situation, including hypotheses about objects or events that are not obviously given and that need to be generated or sought out), and (iii) describing patterns and regularities "beneath the surface of things" (in order to understand, predict, manipulate, and control the modeled situation).

Because of increased recognition about the importance of conceptual models in learning and problem solving, the past decade of research in mathematics and science education has focused on investigating the nature of children's mathematical models. Consequently, in certain areas (such as those related to whole number arithmetic, fractions, and proportions), a great deal is known about the development of instructionally significant models.

For the purposes of this article, it is not necessary to describe details about the preceding models. However, the following trends are relevant to mention because mechanistic conceptions continue to dominate many people's views about the nature of mathematics, teaching, learning, and problem solving, as well as the kind of abilities that are needed for success in an age of information and in a technology-based society.

A comparison of mechanistic and organic views of mathematics, problem solving, learning, and instruction

<table>
<tr><th></th><th>Mechanistic perspectives</th><th>Organic perspectives</th></tr>
<tr><td rowspan="2">The nature of mathematics</td><td>The objectives of instruction are stated in the form: Given ... the student will... That is, knowledge is described using a list of mechanistic condition-action rules (definitions, facts, skills) some of which are higher order metacognitive rules for making decisions about (i) which lower level rules should be used in a particular situation, and (ii) how lower level rules should be stored and retrieved when needed.</td><td>Knowledge is likened not to a machine, but to a living organism. Many of the most important cognitive objectives of mathematics instruction are descriptive or explanatory systems (i.e., mathematical models) used to generate predictions, constructions, or manipulations in real life problem-solving situations or whose underlying patterns can be explored for their own sakes.</td></tr>
<tr><td colspan="2">According to the Mathematical Sciences Education Board's Reshaping School Mathematics: Two outdated assumptions are that: (i) mathematics is a fixed and unchanging body of facts and procedures, and (ii) to do mathematics is to calculate answers to set problems using a specific catalogue of rehearsed techniques. (p. 4). ... As biology is a science of living organisms and physics is a science of matter and energy, so mathematics is a science of patterns. ... Facts, formulas, and information have value only to the extent that they support effective mathematical activity. (p. 12)</td></tr>
<tr><td rowspan="2">The nature of problem solving</td><td>In general, problem solving is described as getting from givens to goals when the path is not obvious. But in mathematics classrooms, problem solving is generally restricted to answering questions that are posed by others, within situations that are described by others, to get from givens to goals that are specified by others, using strings of facts and rules that are restricted in ways that are artificial and unrealistic. In this way, students' responses can be evaluated by making simple comparisons to the responses expected by the authority (the tutor).</td><td>Important aspects of real life problem solving involve developing useful ways to interpret the nature of givens, goals, possible solution paths, and patterns and regularities beneath the surface of things. Solutions typically involve several "modeling cycles" in which descriptions, explanations, and predictions are gradually refined and elaborated. Therefore, several levels and types of responses are nearly always acceptable (depending on purposes and circumstances); students themselves must be able to judge the relative usefulness of alternative models.</td></tr>
<tr><td colspan="2">Problems in textbooks and tests tend to emphasize the ability to create meanings to explain symbolic descriptions; but real problems more often emphasize the ability to create symbolic descriptions to explain (manipulate, predict, or control) meaningful situations. For example, for a mountain climber, the main problem is to understand the terrain of a given mountain or cliff; once the terrain is understood, the activity of getting from the bottom to the top is simply a (strenuous, complex) exercise.</td></tr>
</table>

A comparison of mechanistic and organic views of mathematics, problem solving, learning, and instruction (continued)

	Mechanistic perspectives	**Organic perspectives**
The nature of experts	Humans are characterized as information processors; outstanding students (teachers, experts) are those who flawlessly remember and execute factual and procedural rules–and who are clever at assembling these facts and rules in ritualized settings.	Experts are people who have developed powerful models for constructing, manipulating, and making sense of structurally interesting systems; they are people who are proficient at adapting, and extending, or refining their models to fit new situations.
	In an *age of information* many of the most important "things" that influence peoples' daily lives are communication systems, social systems, economic systems, education systems, and other systems–which are created by humans as a direct result of metaphors that structure the world at the same time they structure humans' interpretations of that world. Therefore, (i) there is no fixed and final state of evolution, even in the context of elementary mathematical ideas, and (ii) reducing the definition of an expert to a single static list of condition-action rules is impossible (in principle), not just difficult (in practice).	
The nature of learning	Learning is viewed as a process of gradually adding, deleting, linking, uncoupling, and debugging mechanistic condition- action rules (definitions, facts, or skills). If the precise state of knowledge is known for an expert (E) and for a given novice (N), then the difference between these two states is portrayed as the subtracted difference (E-N).	Learning involves model building, theory building, and system building; and these constructs develop along dimensions such as concrete-to-abstract, particular-to-general, undifferentiated-to-refined, intuitive-to-analytic-to-axiomatic, situated-to-decontextualized, and fragmented-to-integrated. So evolution involves differentiating, integrating, and refining unstable systems–not simply linking together of stable rules; it involves discontinuities and conceptual reorganizations–such as when students go beyond thinking WITH a given model to also think ABOUT it. Experts not only know more than novices, they also know differently.
The nature of teaching	Teaching involves mainly (i) demonstrating relevant facts, rules, skills, and processes, (ii) monitoring activities in which students repeat and practice the preceding items, and (iii) correcting errors that occur.	Teaching focuses on carefully structuring experiences so students confront the need for mathematically significant constructs, and responses involve constructing, refining, integrating, or extending relevant constructs.
	According to the Mathematical Sciences Education Board's Everyone Counts: *The teaching of mathematics is shifting from an authoritarian model based on "transmission of knowledge" to a student-centered practice featuring "stimulation of learning."* (p. 5) ... *Teachers should be catalysts who help students learn to think for themselves. They should not act solely as trainers whose role is to show the "right way" to solve problems. ... The aim of education is to wean students from their teachers.* (p. 40)	

Again, the point to emphasize about the preceding trends is that when the organic perspectives are emphasized in instruction and assessment, a wider range of students often emerge as knowledgeable and able.

In mathematics, in the past decade an explosion of developments has produced a variety of new types of elementary mathematical systems (such as those based on formerly nonelementary elements involving recursive functions or accumulating quantities); and new levels and types of understandings are also being emphasized for old mathematical systems. For example, even when tools as simple as spreadsheets or graphing calculators are used, their graphics capabilities often provide powerful new structural metaphors that can be explored for their own sakes, or that can be used to create, describe, explain, manipulate, predict, or explore structurally interesting systems in mathematics or the real world.

Again, a significant characteristic of the preceding trend is that models and modeling are assigned roles of central importance. For example, according to the Mathematical Sciences Education Board's publication *On the Shoulders of Giants*, characteristics that distinguish mathematics from other domains of knowledge can be summarized as follows.

> *Mathematics is the science and language of pattern. ... As biology is a science of living organisms and physics is a science of matter and energy, so mathematics is a science of patterns. ...To know mathematics is to investigate and express relationships among patterns: to discern patterns in complex and obscure contexts; to understand and transform relations among patterns; to classify, encode, and describe patterns; to read and write in the language of patterns; and to employ knowledge of patterns for various practical purposes. ... Facts, formulas, and information have value only to the extent that they support effective mathematical activity.* (p.5)

In other words, (i) doing "pure" mathematics means investigating patterns (or systems) for their own sake (by constructing and transforming them in structurally interesting ways and by studying changes in their structural properties), and (ii) doing "applied" mathematics means using patterns (or systems) as models (or structural metaphors) to describe, explain, predict, or control other systems. So again, the ability to construct (and adapt, and use) mathematical models emerges as a central goal of mathematics instruction; and these views are strongly represented in the new and unprecedented national consensus that has been reached about the *Curriculum and Assessment Standards for School Mathematics* from the National Council of Teachers of Mathematics. Yet in studies conducted by Romberg and his colleagues to investigate the alignment of nationally significant standardized tests with the *NCTM Standards,* the conclusions have been consistent.

> *These tests are based on different views of what knowing and learning mathematics means. ... These tests are not appropriate instruments for assessing the content, process, and levels of thinking called for in the STANDARDS.* (Romberg, Wilson, and Khaketla 1991, p. 3)

Mathematical ability does not simply consist of the ability to flawlessly remember and execute intricate sequences of rules. The National Research Council's *Renewing U.S. Mathematics* states,

> *The most important components of mathematical talent cannot be addressed: (i) using timed tests with large numbers of small decontextualized questions, or (ii) when artificial restrictions are placed on the resources that are available. ... Most of the tests used for mathematics assessment have too narrow a focus. They do not measure the wide range of mathematical skills and abilities that educators and business leaders believe is needed for a population to live and work in a world increasingly shaped by mathematics, science, and technology. ... Discontinue use of standardized tests that are misaligned with national standards for curriculum.*

The most important components of what it means to think mathematically involve generating descriptions (or explanations, or manipulations, or constructions, or modifications, or predictions) about the behaviors of structurally interesting systems.

Trends in the "Real World"

In the real world, the essence of an *age of information* is that the most important "objects" that influence peoples' daily lives are human constructs (models, or systems) that are used to mold and shape the world and to make sense of the increasingly complex systems (e.g., communication, transportation, management, political, education, and economic systems) that mental models are used to create.

In ethnographic studies investigating the mathematical capabilities of shoppers, tailors, carpenters, street vendors, and others, it has become clear that most people's "school math" abilities operate relatively independently from their "real math" abilities, and that failure or success in one area does not guarantee failure or success in the other. For example, Resnick has summarized the following reasons why traditional textbooks, teaching, and tests have been inconsistent with real life problem solving and decision making.

- School learning emphasizes individual cognition, while learning in everyday contexts tends to be a cooperative enterprise.
- School learning stresses "pure thought," while the outside world makes heavy use of tool-aided learning.
- School learning emphasizes the manipulation of abstract symbols, while nonschool reasoning is heavily involved with objects and events.
- School learning tends to be generalized, while the learning required for on-the-job competency tends to be situation specific.

Resnick concludes that " ... *school work draws on only a limited aspect of intelligence, ignoring many of the intelligences needed for vocational success, especially in the more prestigious vocations.*" Furthermore, our research suggests that if students are given the opportunity, within problem-solving activities that encourage sense-making based on extensions of students' personal knowledge and experiences, then even students who have been labeled below average often emerge as exceptionally capable by routinely inventing (extending, modifying, or refining) mathematical models that go far beyond those emphasized in their previous academic failures (Lesh and Lamon 1992).

Summary and Conclusions

The foremost issues that have been emphasized throughout this article are that (i) current textbooks, tests, and teaching emphasize narrow, shallow, and noncentral conceptions of mathematical knowledge and ability, and (ii) a promising way to recognize and reward the mathematical potential of a broader range of students is to focus on models and modeling in both instruction and assessment.

In curriculum reform, issues involving technology, content quality, and equity (or fairness to individual students) are often viewed as working at cross purposes. Yet for success to be achieved, progress in these three areas must be mutually supportive. This is why, when we examined trends related to these areas, it is significant to notice that each leads to a greater emphasis on models and modeling. Therefore, an emphasis on models and modeling should provide an ideal path for progress to be made.

In the past, if a narrow conception of talent was correlated with a more representative interpretation, testing specialists tended to treat modest (.5 to .6) correlations as though they were sufficient—that is, sufficient for selecting small-but-adequate numbers of students for access to scarce-but-adequate resources. But today, even beyond concerns about fairness, national assessment priorities have changed. At a national level, our foremost problem is not to *screen* talent; it is to *identify* and *nurture* capable students. The pool of students receiving adequate preparation in mathematics is no longer adequate; and far too many capable students are being shut out or turned off by textbooks, teaching, and tests that give excessive attention to views of mathematics and problem solving that are fundamentally inconsistent with the national *Curriculum and Assessment Standards for School Mathematics.*

REFERENCES

Committee on the Mathematical Sciences: Status and Future Directions, Board on Mathematical Sciences, Commission on Physical Sciences, Mathematics, and Applications, National Research Council. (1990). *Renewing U.S. mathematics: A plan for the 1990s.* Executive Summary. Washington, DC: National Academy Press.

Lesh, R., and Lamon, S. (eds.). (1992). *Assessment of authentic performance in school mathematics.* Washington, DC: American Association for the Advancement of Science.

Mathematical Sciences Education Board, National Research Council. (1989). *Everybody counts - a report to the nation on the future of mathematics education.* Washington DC: National Academy Press.

Mathematical Sciences Education Board, National Research Council. (1990). *Reshaping school mathematics: A philosophy and framework for curriculum.* Washington, DC: National Academy Press.

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

Romberg, T.A., Wilson, L., and Khaketla, M. (1991). *The alignment of six standardized tests with the NCTM Standards.* Madison, WI: University of Wisconsin.

Steen, L.A. (1990). *On the shoulders of giants: New approaches to numeracy.* National Research Council. Washington, DC: National Academy Press.

Enhancing the Success of African American Students in the Sciences: Freshman Year Outcomes*

Kenneth I. Maton and Freeman A. Hrabowski III
University of Maryland Baltimore County

The Meyerhoff Program is an intensive, multicomponent program focused on enhancing the success of talented African American students in science and engineering at a predominantly white, medium-sized university. The program components, taken together, address the four primary factors emphasized in the research literature as limiting minority student performance and persistence in science: knowledge and skills, motivation and support, monitoring and advising, and academic and social integration. Outcome analyses indicated that the first three cohorts of Meyerhoff students (N=69) achieved an overall GPA (mean=3.5) significantly greater than that of a historical comparison sample of African American students (mean=2.8) of comparably talented science students at the university. This difference was even greater for first-year science GPA (means of 3.4 and 2.4, respectively), and in specific science and mathematics courses. Observational and questionnaire data indicated that the Meyerhoff Program study groups, peer-based community, financial scholarships, summer bridge program, and staff appear to be especially important contributors to student success. Implications of the findings for enhancing the success of African American and other underrepresented populations in science are discussed.

Introduction

Although African American students represent 9 percent of those enrolled in universities and colleges, they are awarded only 5 percent of the bachelor's degrees and approximately 2 percent of the doctorates in science and engineering. Furthermore, the number of Ph.D.'s in the physical sciences awarded to African American students decreased from 41 in 1975 to 23 in 1990 (National Research Council 1991). Increasingly, the need for a national effort to enhance the success and persistence of African American (and other minority) students in undergraduate and graduate study in the sciences has been emphasized (Malcom 1991; Pearson and Bechtel 1989). For such a national effort to succeed, key factors contributing to minority student success need to be identified, and model intervention programs incorporating these factors must be developed and evaluated.

The research literature suggests that a number of factors are critical for minority student success, performance, and persistence in science. These include knowledge and skills, motivation and support, monitoring and advising, and academic and social integration (Hrabowski and Maton, in press). Comprehensive intervention programs that effectively address each of the four areas are necessary if the success rates of minority students in science and engineering are to be increased

*Most of the material originally presented and the text of this paper are taken from the article "Enhancing the Success of African-American Students in the Sciences: Freshman Year Outcomes," to be published in an upcoming issue of *School Science and Mathematics*.

substantially. To the extent such programs build upon the strengths of students and mobilize peer, faculty, staff, family, and community resources to meet minority student needs, they should substantially increase minority student success in science. The Meyerhoff Program, named after the philanthropist who provided the initial funding, at the University of Maryland Baltimore County (UMBC) is an intervention program designed to accomplish that outcome.

The Meyerhoff Program

The Meyerhoff Program was created to address the key student needs and related environmental factors emphasized in the research literature. The program is based on a strengths model, which assumes that all students selected are capable of succeeding in science, given the proper resources and opportunities. The 13 program components are as follows:

1. **Recruitment**. Top math and science students are sought to participate in an on-campus selection weekend involving faculty, staff, and peers.
2. **Summer bridge program**. Services include math, science, and humanities coursework, training in analytic problem solving, group study, and social and cultural events.
3. **Scholarship support**. Meyerhoff scholars receive 4-year, comprehensive scholarships; finalists receive more limited support. Continued support is contingent on the student maintaining a B average and remaining a science or engineering major.
4. **Study groups**. Group studying among students is strongly and consistently encouraged by program staff.
5. **Program values**. The program values, which are consistently emphasized, include striving for outstanding academic achievement, seeking help (tutoring, advisement counseling) from a variety of sources, supporting one's peers, and preparing for graduate or professional school.
6. **Program community**. The program represents a family-like, campus-based community for students. Staff hold group meetings with students regularly; students live in the same residence halls during their freshman year.
7. **Personal advising and counseling**. A full-time program academic advisor, along with the program executive director, director, and assistant, regularly monitor and advise students.
8. **Tutoring**. All Meyerhoff students are encouraged to take advantage of department and university tutoring resources in order to optimize course performance.
9. **Summer research internships**. Program staff use a network of contacts to arrange summer science and engineering internships.
10. **Faculty involvement**. Department chairs and faculty are involved in all aspects of the program, including recruitment, teaching, research mentorship, and special events and activities.
11. **Administrative involvement and public support**. The program receives high-level campus administrative support and public support.
12. **Mentors**. Each student has a mentor recruited from among Baltimore-Washington area professionals in science, engineering, and health.
13. **Family involvement**. Parents are kept informed of student progress, invited to special counseling sessions as problems emerge, included in various special events, and joined together in a mutual support resource, the Meyerhoff Family Association.

The Meyerhoff Program components, taken together, address each of the four primary areas of minority student need noted earlier. Importantly, they do so by mobilizing and building upon each of five major sources of influence upon students: peers, faculty, administrative staff, family, and community members. Table 1 presents an overview of the different areas of student need as affected by the various program components.

The Current Study

The purposes of the current research are to analyze outcome data concerning the initial academic impact of the Meyerhoff Program and to examine process data concerning factors likely responsible for program impact. Academic impact is examined by comparing the freshman year academic outcomes of the first three cohorts (N=69) of Meyerhoff students with those of equally talented African American students in the sciences and engineering who attended UMBC prior to the advent of the program. Student experiences were examined using both direct observation of various aspects of student academic life and student questionnaire responses.

Table 1.
Meyerhoff Program components and minority student needs

Meyerhoff program components	Minority student needs in math/science			
	Knowledge and skills	Motivation and support	Monitoring and advising	Academic and social integration
Recruitment............................. .	X[1,3]	X[1,2,3]		
Summer bridge program	X[2,3]	X[1,2,3]	X[3]	X[1,3]
Scholarship.............................. .		X[3]		
Study groups	X[1]	X[1]		X[1]
Program values	X[3]	X[1,3]	X[3]	X[1,2]
Program community.................. .		X[1,3]	X[1,3]	X[1,3]
Personal advising and counseling ..		X[3]	X[3]	X[3]
Tutoring...............................	X[3]			
Summer internship	X[5]	X[5]		
Faculty involvement	X[2]	X[2]		X[2]
Administrative involvement and support		X[3]		X[3]
Family involvement.................. .		X[3,4]	X[3,4]	
Mentors		X[5]	X[5]	

1 Program component mobilizes and builds upon peer support.
2 Program component mobilizes and builds upon faculty support.
3 Program component mobilizes and builds upon staff support.
4 Program component mobilizes and builds upon family support.
5 Program component mobilizes and builds upon community support.

Method

Research Participants

The primary sample includes 69 Meyerhoff Program students and 43 comparison students. The Meyerhoff students are from the first three program cohorts (19 in the class entering in 1989, 15 entering in 1990, and 35 in 1991). Two Meyerhoff students who were added to the second year cohort after the program had already started are not included in the primary sample.

The 43 comparison students are African American students who attended UMBC. All but two were freshmen before the program was initiated, and those two entered UMBC the initial year of the Meyerhoff program (after that first year, many non-Meyerhoff African American students became involved to varying extents in program activities). All had academic records at or above the minimum cutoffs used to select Meyerhoff students—Math SAT of 550, combined SAT of 1,050, and high school GPA above 3.0; or Math SAT above 500, combined SAT above 1,000, and high school GPA above 3.7 (most Meyerhoff and comparison students had SAT and GPA scores substantially above these minimum cutoffs). Finally, all were likely candidates for a career in science, as indicated either by a declared science major or freshman year coursework that included at least one math or science course required for a science major (for the purposes of this study, science encompasses chemistry, physics, biology, engineering, math, and computer science). Of the 43 comparison students, 26 entered UMBC between 1985 and 1989, 11 between 1980 and 1984, and 6 before 1980.

Procedure

Student academic and demographic data for the Meyerhoff and historical African American comparison samples were obtained from the institutional records at UMBC. Information about current student experiences for a subsample of 30 Meyerhoff freshmen (from the 1990-91 class) was collected in the spring of 1991 from direct observation and questionnaire responses. Permission to collect the data was obtained from the sampled students before the study began.

Two African American graduate students were used to gather this information. Observation of each student was carried out in class, study groups, program meetings, or tutoring or advising sessions. Extensive field notes on the supportive and stressful aspects of each setting observed were recorded following the observation; additionally, students were asked to complete a questionnaire that included items focused on their perception of the supportiveness of various program and university resources. Comparable observational data were obtained concurrently from 25 African American freshmen in the sciences with strong academic backgrounds (though generally not at the level as the Meyerhoffs) who were taking one or more of the same math, science, or engineering classes that the Meyerhoff students were taking.

There were some differences in student characteristics between the full Meyerhoff sample and the historical comparison sample, which are detailed in the Results section of this report. In order to compensate for these differences, a matching process was used to select from the full sample a subgroup of students who were more equivalent. One-to-one matching was based on gender, SAT scores, and high school GPA. The subsample contained 34 Meyerhoff and 34 comparison students. All analyses were carried out on both the matched and full samples.

Results

Preliminary Analyses

Table 2 presents student background characteristics for the matched and full study samples. The left-hand section of the table indicates that the matching process successfully created equivalent samples of Meyerhoff and comparison students. Specifically, the matched groups are comparable in terms of SAT-Math scores (means of 617.1 and 613.8), SAT-Verbal scores (means of 540.6 and 539.6), high school GPA (means of 3.48 and 3.43), and gender (55.9 percent female).

The right-hand section of Table 2 indicates that the full Meyerhoff sample had significantly higher SAT-Math scores (mean=635.7) than the comparison sample (mean=607.7), $\underline{t}$=2.64, $\underline{p} < .01$. Although the Meyerhoff students had somewhat higher SAT-Verbal scores (mean=544.9) than comparison students (529.7), and slightly higher high school GPAs (3.52) than comparisons (3.41), the two groups did not differ significantly on these variables, $\underline{t}$=1.29, ns, and $\underline{t}$=1.78, ns, respectively. However, there was a significantly lower percentage of females in the Meyerhoff sample (36.2 percent) than in the comparison sample (65.1 percent), χ^2 (1)= 8.87, $\underline{p} < .01$, primarily because the first Meyerhoff cohort consisted entirely of males.

Table 2.
Background characteristic means (and standard deviations) for Meyerhoff and comparison students: Matched and full samples

Student characteristic	Student background characteristic			
	Matched Meyerhoff (N=34)	Matched comparison (N=34)	Full Meyerhoff (N=69)	Full comparison (N=43)
SAT-Math............................. .	617.1 (44.3)	613.8 (48.4)	635.7** (59.0)	607.7 (46.3)
SAT-Verbal............................ .	540.6 (49.5)	539.6 (67.8)	544.9 (55.7)	529.7 (68.5)
High school GPA................. .	3.48 (0.30)	3.43 (0.33)	3.52 (0.33)	3.41 (0.32)
Percent female	55.9%	55.9%	36.2%**	65.1%
Number of science course credits freshman year[1]	20.0* (4.8)	16.7 (6.1)	20.6*** (4.4)	15.2 (6.3)

* $\underline{p} < .05$.; ** $\underline{p} < .01$; *** $\underline{p} < .001$.

[1]All courses in chemistry, physics, biology, engineering, math, and computer science taken during the academic year and the summer were included.

First Year Academic Outcomes

Analyses of Covariance (ANCOVAs) were performed on overall freshman year GPA, science GPA, nonscience GPA, and four key courses necessary for majors in science and engineering—Calculus, Principles of Chemistry, Introductory Engineering, and Concepts of Biology. The covariates included SAT-Math, SAT-Verbal, high school GPA, gender (for the full sample only), year of college entrance, and total number of science credits taken in the freshman year.

Table 3 presents the correlations between background characteristics and freshman year science GPA. Whereas high school GPA is an important predictor of first year science GPA for comparison students, it is not for the Meyerhoff students.

Table 4 indicates the ANCOVA results for both the matched and the full samples. Actual and adjusted means are reported. For the matched samples, four of seven analyses were significant. Meyerhoff students achieved a higher overall GPA (actual mean=3.4) than comparison students (actual mean=2.8), $F(1,61)=3.47$, $p < .001$. Meyerhoff students also achieved a higher science GPA (mean=3.4) than comparison students (mean=2.5), $F(1,61)=6.77$, $p < .001$. Concerning specific courses, among students taking Calculus, Meyerhoff students (N=33) achieved strikingly higher grades (mean=3.5) than comparison students (N=25; mean=2.2), $F(1,50)=14.62$, $p < .001$. In addition, Meyerhoff students taking Principles of Chemistry (N=21) achieved much higher grades (mean=3.4) than comparison students (N=17; mean=2.8), $F(1,43)=6.11$, $p < .05$. There were no significant differences between groups on nonscience GPA (means of 3.5 and 3.2, respectively), Introductory Engineering (means of 3.5 and 3.1), and Concepts of Biology (means of 3.1 and 3.1).

When analyses were repeated using the full samples (with gender included as a sixth covariate), the findings were essentially identical to those with the matched sample, with one exception. The difference between the Introductory Engineering course grades of the Meyerhoff students (N=30; actual mean=3.6) and the comparison students (N=9; actual mean=2.8) achieved statistical significance, $F(1,31)=4.71$, $p < .01$.

Table 3.
Correlations of background characteristics and science GPA[1] for Meyerhoff and comparison samples: separately and combined

Student characteristic	Matched sample			Full sample		
	Meyerhoff	Comparison	Combined	Meyerhoff	Comparison	Combined
SAT-Math.......................	15	-.24	-.08	.15	-.15	-.08
SAT-Verbal.....................	-.23	.29	.13	-.11	.32*	.13
High school GPA...........	.12	.57**	.39**	.15	.40**	.39**
Percent female	.15	.23	.16	.02	.06	.16
Year of college entry.........	.02	-.22	.30*	.13	-.11	.30*
Science course credits freshman year[1]	.20	.41**	.43**	.25*	.38*	.43**
Group (Meyerhoff)...........	--	--	.55**	--	--	.61**

* $p < .05$; ** $p < .01$.

[1]All courses in chemistry, physics, biology, engineering, math, and computer science taken during the academic year and the summer were included in the science GPA.

Table 4.
First year college outcomes for Meyerhoff and comparison students: Matched and full samples[1]

Outcome	Matched Meyerhoff	Matched comparison	Full Meyerhoff	Full comparison
Overall GPA	3.4***	2.8	3.5***	2.8
	(3.5)	(2.8)	(3.5)	(2.8)
Science GPA[2]	3.4***	2.5	3.4***	2.4
	(3.4)	(2.5)	(3.4)	(2.5)
Nonscience GPA	3.5	3.2	3.5	3.2
	(3.6)	(3.2)	(3.5)	(3.2)
MATH 151: Calculus & Analytic Geometry	3.5***	2.2	3.6***	2.3
	(3.6)	(2.1)	(3.5)	(2.3)
CHEM 101: Principles of Chemistry	3.4*	2.8	3.4**	2.7
	(3.6)	(2.7)	(3.5)	(2.7)
ENES 101: Introductory Engineering Science	3.5	3.1	3.6**	2.8
	(4.0)	(2.6)	(4.0)	(2.3)
BIOL 100: Concepts of Biology	3.1	3.1	3.2	2.9
	(3.6)	(2.6)	(3.4)	(2.6)

* $p < .05$; ** $p < .01$; *** $p < .001$.

[1]Statistical tests employed were analyses of covariance with the following covariates: gender (full sample only), SAT-M, SAT-V, high school GPA, year of college entrance, and number of science credits first year. For specific courses, the sample size ranged from 20 to 93, depending on the number of students in the subsample who had taken the particular course.

[2]All courses in chemistry, physics, biology, engineering, math, and computer science taken during the academic year and the summer were included in the science GPA.

NOTE: The adjusted means are shown in parentheses below the actual means.

Clearly the Meyerhoff students are achieving at a substantially higher level than the comparison students, especially in key freshman year math, science, and engineering courses. These differences are striking in magnitude and are present for both the matched and full samples, with key background variables controlled.

Process Data

The outcome results suggest that the Meyerhoff program is making a difference in the freshman year academic achievement of talented African American students in the sciences. Direct observation of Meyerhoff and non-Meyerhoff students' study groups, advising sessions, and program meetings provide insight into the components of the Meyerhoff Program likely related to academic success. For example, observation of Meyerhoff student study groups indicated a strongly supportive, challenging, and task-focused study environment in which students both provided and received help, as necessary. In contrast, the study groups of non-Meyerhoff students taking the same courses had notably less task focus and intensity. The greater intensity and task focus of the Meyerhoff study groups may reflect greater time urgency due to a high level of academic and program engagements and a strong motivation to achieve high grades because of high program expectations and the students' desire to retain their scholarship aid.

Observation of advising sessions indicated fundamental advantages enjoyed by the Meyerhoff students. All students see the same full-time Meyerhoff Program advisor, who is an African American. At least three students are scheduled at a time to help achieve a primary strategic goal of the program—that no student is in a section of a course alone. Group discussion focuses both on individual problems and career decisions. The group advising appears to create an opportunity for students to learn from others' mistakes, to encourage each other, and to bond with each other.

Observation of Meyerhoff Program meetings is especially instructive. In these meetings students regularly are asked to share both their accomplishments and their personal or academic problems—thus, everyone knows how everyone else is doing. The students are regularly and periodically challenged by program staff to perform at high levels, to seek out help from all available sources, and to support each other. Furthermore, they are repeatedly told they are special, that each one of them is fully capable of outstanding performance, and that each one has a special contribution to make in college, in graduate or professional school, and in their future occupations, as well as an obligation to "give back" to the larger African American community.

Concerning questionnaire results, Meyerhoff students completed items focused on various freshman year support sources (5-point scales with 1=low support, 5=high support). Five sources received especially high ratings: study groups (mean=4.4), being part of the Meyerhoff Program community (4.4), financial scholarships (4.3), summer bridge program (4.3), and Meyerhoff staff (4.3) (Table 5). Table 5 also

Table 5.
Perceptions of support and academic performance among 1991-92 Meyerhoff Program freshmen[1]

Support source	Means and percent endorsements on support sources scale items					
	Mean	Not at all supportive		Moderately supportive		Extremely supportive
		1	2	3	4	5
Study groups for science, math, and engineering courses.................... .	4.40	0.0%	0.0%	10.0%	40.0%	50.0%
Being part of Meyerhoff community	4.37	0.0	6.7	10.0	23.3	60.0
Financial scholarships................ .	4.33	3.3	0.0	13.3	26.7	56.7
Summer bridge program............. .	4.30	0.0	0.0	20.0	30.0	50.0
Meyerhoff staff.......................... .	4.27	0.0	0.0	16.7	40.0	43.3
Family involvement.................. .	4.13	3.3	6.7	16.7	20.0	53.3
Academic tutoring..................... .	3.75	7.4	7.4	22.2	29.6	33.3

Percent endorsements and means on academic performance scale items

1. Overall, how would you say you are doing in your field of study?

Not at all	Moderate Extent	Large Extent	Mean
1	2	3	
0.0%	33.3%	66.7%	2.67

2. To what extent do you think you are doing well because of the Meyerhoff Program support?

Not at all	Moderate Extent	Large Extent	Mean
1	2	3	
0.0%	50.0%	50.0%	2.50

[1]Of 35 1991-92 Meyerhoff freshmen, 30 completed questionnaires.

indicates that Meyerhoff students attribute their academic success either to a moderate or large extent to the Meyerhoff Program. Additional questionnaire findings (Table 6) indicate that Meyerhoff students report lower levels of academic stress on many variables than non-Meyerhoff African American freshmen (secondary comparison sample) in the sciences.

In summary, the process evaluation data suggest that, taken together, the financial support, study groups, program community, specialized advising, high expectations, and peer solidarity provide a program environment highly supportive of strong academic performance. Additional outsider observations of a number of these facets of the Meyerhoff program are presented in a recent article (Gibbons 1992).

Table 6.
Indicators of stress[1] among 1991-92 Meyerhoff and 1991-92 African American non-Meyerhoff students[2]

Indicator	1991-92 Meyerhoff	1991-92 non-Meyerhoff	p
Excessive academic load	2.7	3.4	.05
Preparation for the job market	2.4	3.0	.05
Poor study skills and habits	2.2	2.9	.05
Answering essay questions	2.1	3.0	.01
Identity issues	2.0	1.4	.05
Attitudes of administrators and staff to African American students' needs	1.9	2.5	.05
Financial difficulties	1.8	2.8	.01
Lack of financial support	1.7	2.7	.01
Attitudes of faculty toward African American students	1.7	2.3	.05
Poor academic advising	1.6	2.3	.05
Low academic expectations of faculty for African American students' performance	1.5	2.3	.05

[1] 5-point Likert-type scale was used with following descriptors: 1=no stress; 3=average stress; 5=tremendous stress.
[2] 30 Meyerhoff and 25 non-Meyerhoff students completed questionnaires.

Discussion

The results indicate that African American science and engineering students enrolled in a multicomponent intervention program achieve superior freshman year academic outcomes than equally academically talented African American students in a historical comparison sample. The research literature suggests that many factors influence the academic achievement of African American students, including knowledge and skills, motivation and support, monitoring and advising, and academic and social integration. The results of the current research suggest that intervention programs that focus explicitly on each of these factors have the potential to substantially enhance the academic performance of science students in their critical first year of college. Furthermore, the relatively mediocre performance of the comparison sample indicates that without such a multifaceted program, even very talented African American students are unlikely to achieve the levels of academic success necessary to substantially increase the number of Ph.D.'s awarded to students in the physical sciences.

The current research cannot completely rule out alternative explanations of findings (e.g., historical changes in the university or in the preparation of students in the sciences; sample selection biases). However, in our view they appear unlikely to explain fully the dramatic differences in freshman year achievement between the two groups. The combination of outcome and process findings, and the fact that academic background variables were controlled (statistically and through matching), strongly suggest that the Meyerhoff Program is having a substantive impact on the success of African American students in science. Research is underway to obtain additional comparison samples to examine various alternative explanations of findings and to examine the extended outcomes of the program, including both college completion and graduate education through the Ph.D.

REFERENCES

Gibbons, A. (1992). Minority programs that get high marks. *Science*, 258, 1190-1196.

Hrabowski, F.A. III, and Maton, K.I. In press. Enhancing the success of African-American students in the sciences: Freshman year outcomes. *School Science and Mathematics*.

Malcom, S.M. (1991). Editorial. *Science*, 254, 353.

National Research Council. (1991). *Doctorate recipients from United States universities*. Washington, DC: National Academy Press.

Pearson, W., Jr., and Bechtel, H.K. (eds.) (1989). *Blacks, science, and American education*. New Brunswick, NJ: Rutgers University Press.