DOCUMENT RESUME

ED 472 338 SE 067 136

AUTHOR	Millar, Susan B., Ed.
TITLE	Indicators of Success in Postsecondary SMET Education: Shapes of the Future. Synthesis and Proceedings of the Annual NISE Forum (3rd, February 23-24, 1998). Workshop Report.
INSTITUTION	National Inst. for Science Education, Madison, WI.
SPONS AGENCY	National Science Foundation, Arlington, VA.
REPORT NO	NISE-R-6
PUB DATE	1998-11-00
NOTE	163p.
AVAILABLE FROM	National Institute for Science Education, University of Wisconsin-Madison, 1025 W. Johnson Street, Madison, WI 53706. Tel: 608-263-9250; Fax: 608-262-7428; e-mail: niseinfo@macc.wisc.edu. For full text: http://www.wcer.wisc.edu/ nise/Publications.
PUB TYPE	Collected Works - Proceedings (021)
EDRS PRICE	EDRS Price MF01/PC07 Plus Postage.
DESCRIPTORS	Curriculum; Educational Change; *Evaluation; Higher Education; Mathematics Education; Science Education; Teaching Methods; Technology Education

ABSTRACT

This document presents the written records of the Third Annual National Institute for Science Education (NISE) Forum on indicators of success in postsecondary science, mathematics, engineering, and technology (SMET) education. The primary goal of this Forum was to initiate a national dialogue about how assessment and evaluation are and should be used to foster improvements in SMET education at all levels in the U.S. higher education systems. The records, all of which appear in this Proceedings, comprise the opening keynote, a digest of the three panel discussions, the remarks of panel discussants, a synthesis of participants' observations written after each panel discussion (think pieces), the closing reflections, an analysis of the participants' theories of change, and the presenters' papers. A list of acronyms and a list of the approximately 300 participants' names and location information are provided in the appendix. The panelists papers include: (1) "Assessment as a Learning Process: What Evidence Will We Accept That Students Have Learned?" (Diane Ebert-May); (2) "Moving the Mountain: Impediments to Change" (Eric Mazur); (3) "The Integrality of Assessment" (David B. Porter); (4) "The Reaction to the Symptoms Versus the Reaction to the Disease" (Brian P. Coppola); (5) "Assessment and the Promotion of Change in Community Colleges" (Eileen L. Lewis); (6) "Assessing and Evaluating the Evaluation Tool: The Standardized Test" (Richard Tapia); (7) "Assessment, Evaluation and Accreditation: Are We on the Same Wave Length? Or How Does One Provide Linkages for Systemic Change?" (Jack Bristol); (8) "An Assessment Model to Drive Undergraduate Educational Reform in the SMET Fields in a Large Public Multicampus University System" (Manuel Gomez); and (9) "Technology-Assisted Learning in Higher Education: It Requires New Thinking about Assessment" (Sheri D. Sheppard). (MM)

National Institute for Science Education

University of Wisconsin-Madison . National Center for Improving Science Education

Workshop Report No. 6

Indicators of Success in Postsecondary SMET Education: Shapes of the Future

Synthesis and Proceedings of the Third Annual NISE Forum

Edited by Susan B. Millar

Funded by the
National Science Foundation

Workshop Report No. 6

Indicators of Success in Postsecondary SMET Education: Shapes of the Future

Synthesis and Proceedings of the Third Annual NISE Forum

Edited by Susan B. Millar

National Institute for Science Education
University of Wisconsin–Madison

November 1998

Acknowledgments

The authors extend their thanks to Ann Burgess, Susan Daffinrud, Arthur Ellis, Terrence Millar, Joel Mintzes, Andrew Porter, and Paula White for their thoughtful review of the draft version of this document. We also thank Debbie Stewart and Ingrid Rosemeyer for their good suggestions and patience during the editing process.

Contents

Introduction: A Framework for Forum Contributions

Elaine Seymour

This document presents the written records of the Third Annual National Institute for Science Education (NISE) Forum on indicators of success in postsecondary science, mathematics, engineering and technology (SMET) education. The records, all of which appear in this Proceedings, comprise the opening keynote, a digest of the three panel discussions, the remarks of panel discussants, a synthesis of participants' observations written after each panel discussion (think pieces), the closing reflections, an analysis of the participants' theories of change, and the presenters' papers. A list of acronyms and a list of the approximately 300 participants' names and location information are provided in the appendix.

Arthur Ellis, Elaine Seymour, and Susan Millar led the Forum development team. This core group was assisted by program officers from the National Science Foundation's (NSF) Division of Undergraduate Education and by the following members of NISE's College Level One team: Aaron Brower, Ann Burgess, Anthony Jacob, Kate Loftus, Robert Mathieu, and Catherine Middlecamp.

Forum Goal and Design

The primary goal of this Forum was to initiate a national dialogue about how assessment and evaluation are and should be used to foster improvements in SMET education at all levels in the U.S. higher education system. To focus attention on different levels within the system, three different but related topics—assessment at the classroom level; assessing learning as an aspect of change in classrooms, disciplines, and institutions; and the role of evaluation in change at the department, institutional, and national levels—were chosen. The following strategies were used to foster productive and genuine dialogue about these topics:

- Panelists and panel facilitators were selected on the basis of their rich experiences with assessment and evaluation and their diverse backgrounds within higher education. The panel facilitators were Brock Spencer, professor of chemistry and principal investigator of the NSF-funded ChemLinks Chemistry Systemic Reform project, Beloit College; Elaine Seymour, director of the Ethnography and Evaluation Research Center at the University of Colorado at Boulder; and Clifford Adelman, Senior Research Analyst at the Office of Educational Research and Improvement, U.S. Department of Education.
- Scientists, mathematicians, engineers and administrators from many disciplines were invited by contacting members of higher education organizations such as the American Association of Higher Education (AAHE), the American Geological Institute, the American Chemical Society, and the American Physical Society.
- Panelists were asked to write short papers for participants and the other panelists to read ahead of time.
- The facilitators worked with the panelists to plan a "fishbowl" format that allowed them to explore a set of linked themes developed in their various papers, and also to engage the audience in the discussion.
- Officers of the NSF were asked to frame the discussion, with the agency's director, Neal Lane, and its Assistant Director for Education and Human

Resources, Luther Williams, opening the Forum with a plenary address, and with Norman Fortenberry (Director, Division of Undergraduate Education), Daryl Chubin (Director, Division of Research, Evaluation, and Communication), and Larry Suter, (Deputy Director, Division of Research, Evaluation and Communication)

- Two institutional administrators, Cora Marrett, Provost, University of Massachusetts-Amherst, and John Wiley, Provost, University of Wisconsin-Madison, were chosen to provide a real-time synthesis and closing remarks.
- Participants were assigned to discussion groups that met after each panel session. They were asked to spend the first fifteen minutes responding to the panel in written think pieces that were used to produce the *Synthesis* portion of this document. The remainder of each discussion group session was spent sharing knowledge and views.
- Participants were invited to present poster sessions for perusal by others during a lengthy lunch period.

A second goal of this Forum was to inform the NISE's College Level One team of the needs of the postsecondary SMET education community with respect to assessment. This information is being used by the College Level One team's 1998-99 Institute on Assessment in College Level SMET Classrooms. For general information about this team and its work following on this February 1998 Forum, readers may wish to consult the NISE website (*http://www.wcer.wisc.edu/nise*), and the College Level One team's website (*http://www.wcer.wisc.edu/nise/CL1/*).

Guide to the Reader

This document is intended primarily to provide the participants, officers of the NSF, members of the NISE, and the national community with a thematic account of the ideas and issues developed for and during this Forum. Each section of this document considers, from a different perspective, the major issues raised, courses of action undertaken or advocated, areas of agreement and dissent, expectations, and unmet needs. A few specific comments about the *Panel Discussion Summaries and Commentaries*, the *Synthesis,* and the *Analysis of Participants' Theories of Change* sections are provided to guide the reader.

The *Panel Discussion and Summaries* section is designed for readers seeking an understanding of the key points made during the panel sessions. Each of the three summaries presents the key themes and points made by the panelists. The *Commentary* following each summary is the full transcript of the remarks made by the panel discussant. As the panelists' remarks were made on the assumption that the participants had read the panelists' papers, readers may find it useful to read the relevant papers (see Appendix A) in conjunction with the summary of each panel.

The *Synthesis of Participants' Think Piece Essays* section is designed for readers with a special interest in understanding the views of the assembled 300 forum participants. The majority of Forum participants were faculty and administrators from four-year colleges and universities. A number of representatives from two-year colleges, K–12 educational systems, national agencies, and professional organizations were also in attendance. This section synthesizes the salient points appearing in the nearly 600 different think pieces written by the participants after the three panel sessions, attempting to include points made by participants from each of these stakeholder groups. As the think piece essays were written in response to, and frequently referred to specific points made in, the

panel discussions, those who read both the panel summaries and the synthesis will note that forum participants reinforced many of the same points made by the panelists, but also raised important issues not addressed by the panelists.

The *Analysis of Participant Theories of Change* offers a broad conceptual framework for the key points made in all the components appearing in this Proceedings, and integrates themes developed by all Forum participants. Readers with an interest in the *process* of higher education reform will find this piece of special interest.

Main Themes and Voices from the Forum

The Centrality of Assessment in Higher Education Reform

That a forum focused on assessment should be convened at this time reflects a particular stage in the nationwide conversation among postsecondary educators about the current and future expectations and needs of higher education and the role of SMET education within it. Assessment is an appropriate focus for a discussion of next steps because, as our collective experience with reform issues has grown, we have come to understand that issues of measurement and evaluation are, as panelist Brian Coppola observed, actually or potentially "linked to every other aspect of the instructional setting."

The Forum participants identified the following functions of assessment as critical in the endeavor to enhance the quality of SMET higher education:

- exploration and clarification of student learning gains in conceptual understanding and in the ability to make connections, formulate research questions, and communicate knowledge; and
- evaluation of teaching in ways that more accurately reflect its efficacy in enabling learning.

Many participants identified as new to the higher education SMET culture a growing realization that these two types of assessment—of students and of faculty—should focus on learning and are essentially linked. Participants proposed that assessment data from these two sources should be used:

- to provide useful feedback to both teachers and students on their work;
- to establish the worth of reformed teaching activities;
- as a basis for argument and persuasion to promote further improvements in the quality of SMET education;
- to protect from negative career consequences faculty who work in contexts that are unsympathetic to their classroom innovations; and
- as primary resources for institutions, disciplines, and national agencies that need data gathered at classroom, departmental, and institutional levels to develop aggregate measures of progress.

The Search for Indicators at All System Levels

Forum participants were hopeful that assessment could provide coherent, workable, and cost-effective indices of progress and accountability for individual faculty, departments, and whole institutions. Such indices must be grounded in more accurate and transferable measurement of student learning gains. They also must be capable of assessing the effectiveness with which the system is "making provision" for SMET education that is of high quality yet affordable (Luther Williams, opening keynote).

The Forum participants agreed that the process of developing cost-effective, accurate, and transferable indicators of system quality is

challenging. Larry Suter (discussant) observed that he has been actively addressing this challenge since 1990, when he began turning more attention to the national data about the undergraduate sector. He found that he "could not report how much our students were learning, how the pipeline was working, or where the holes were. . . . There were lots of statistics. . . . What was missing was the organization of information around subject areas." Clifford Adelman (panel facilitator) agreed that "without some metrics for aggregation, all you have are fragments." In addressing the challenge to improve these data, Suter said that the NSF has experienced difficulty finding people "who are willing to tackle the measurement problem" facing postsecondary education. He observed that in the development of measurements, there is not yet "the same kind of leadership in the higher education area as there is in elementary and secondary education." However, Suter was hopeful that a number of recently funded projects would develop national indicators that give a better public account of higher education teaching and learning. In addition, he hoped that the Forum would encourage more people to become engaged in discussions about measurement.

In Williams' view, "the two domains" of student learning gains and system indicators "interconnect, but one does not substitute for the other." Other contributors were hopeful that common measures of student learning used across departments could collectively provide systemic indicators of what Manuel Gómez (panelist) called the institution's educational health. The ongoing task for education scholars, assessment and evaluation specialists, funding agencies, and the reform community is to grapple with the difficulties of how to develop measurements that will serve every level of the system and the system overall.

Cross-Currents: The Viewpoints, Expectations, and Needs of Participants

The papers, think pieces, panel discussions, and observations from the floor indicated several subsets of interest in assessment and differences in the nature of the participants' engagement with assessment and with curriculum reform issues. The viewpoints outlined below reflect the wide invitational character of the Forum, and differences in the nature of the participants' engagement with issues of educational quality and change.

Deeply Engaged Voices

Some views reflected deep and long engagement with the improvement of SMET higher education and with the standards by which it is judged. These observers took it as a given that changes were underway. Many were highly engaged in reform activities as teachers researchers, and administrators. Some looked to assessment as the means to inform, improve, and validate their own work and to move the reform movement forward. Some expressed regret that the Forum did not address in a more systematic way what is already known about assessment and evaluation methods applied specifically within higher education. They strongly debated various strategies for classroom-level assessment and for leveraging change at the departmental and institutional levels.

Newly Engaged Voices

The Forum also included faculty who had more recently discovered what panelist Diane Ebert-May and others referred to as "active learning." Their expressed expectation was that the Forum would offer practical guidance in how to better design and use assessments. While they gained knowledge of tools and instruments that they could adapt for their own classroom use from the panel and poster sessions, they also looked for more. They used the small-group discussions and informal encounters as opportunities for networking an

exchange of assessment materials and methods. This clearly expressed need speaks to the recommendations made by a number of speakers that national and institutional leadership should increase access to workshops on curriculum development, pedagogical techniques, and the design of learning assessments.

Gradualist Voices

A third set of participants was less concerned with "change" or "reform" than with assessment in the service of enhanced cur

riculum quality. As John Wiley (one of the closing speakers) emphasized, better assessment techniques would allow clearer judgement of the relative merits of different approaches to teaching—whether more or less traditional. These gradualist voices also reminded the Forum participants that large and important sections of the SMET community remain skeptical about the need for curriculum reform or for a new emphasis on learning, and that a temperate choice of language in advocating these ideas may be wise. As William Clune (participant) wrote, "We're talking about careful change."

Community College and K–12 Voices

Participants whose comments reflect the experience of SMET teachers in community colleges and the K–12 system urged reformers to take into account the needs of their students and to learn from K–12 teachers and community college instructors, who have much longer experience with curriculum reform and professional development. They pressed for the alignment of the new goals and practices in four-year higher education institutions with those of the two-year college systems. Eileen Lewis (panelist) pointed out that the prevailing alignment may not be in the desired direction. The prestige of research institutions is such that their pedagogical strategies exert a conservative influence on teaching methods in community colleges.

Voices Calling for the Evaluation of Evaluation Criteria

The last voice, heard from across all groups, expressed dissatisfaction with the following common assessment practices:

- Use of standardized tests (such as the SAT) as an "inappropriate filter" that effectively deprives institutions of talented students who have poorer access to good college preparation. Panelist Richard Tapia argued that the effect of this practice is to reinforce class and race/ethnic bias in access to quality higher education.
- Use of institutional assessment instruments whose purpose is obscure and methods are faulty. Participants were concerned that the use of poorly designed instruments has left faculty indifferent or suspicious of assessment in general and, thus, disinclined to consider the use of institutional assessment practices, regardless of their quality.
- Use of poorly designed end-of-semester course evaluation instruments. Faculty reported that these instruments commonly fail to give information about student learning gains, offer poor feedback (except in the students' write-in comments), and sometimes are used inappropriately in tenure or promotion decisions.

All of these commentators' objections to institutional assessment focused on the poor quality of many of the instruments they are obliged to use. Participants called for re-examination of the criteria on which commonly used assessment instruments are based and evaluation of the consequences of using these instruments. As Tapia argued, "The criteria they reflect can hinder outcomes that we value."

The reader will find these themes and these voices recurring throughout the document that follows. Together, they reflect the stage that we have reached in our common search for improved quality, and for indicators thereof, in SMET higher education.

Forum Agenda

Third Annual NISE Forum: February 23-24, 1998

Indicators of Success in Postsecondary Science, Mathematics, Engineering and Technology (SMET) Education: Shapes of the Future

Monday February 23

Opening Session

8:30-9:15 **Welcome, Overview**
Andrew Porter, National Institute for Science Education
Neal Lane, National Science Foundation

Shaping the Future
Luther Williams, National Science Foundation

Panel 1: Assessment of Teaching, Learning, and Curriculum Change in SMET Classrooms

9:15-10:45 **A. Panel Discussion**
Panel Chair: Arthur Ellis, National Institute for Science Education
Panel Facilitator: Brock Spencer, Beloit College
Panel Members: Diane Ebert-May, Northern Arizona University
Eric Mazur, Harvard University
David B. Porter, U.S. Air Force Academy
Commentary: Norman L. Fortenberry, National Science Foundation

10:45-11:15 *Break*

11:15-12:30 **B. Small Group Discussions**

12:30-2:15 *Lunch*

C. Poster Session

Panel 2: Assessment and the Promotion of Change in Departments, Disciplines, and Institutions

2:15-4:00 **A. Panel Disscussion**

Panel Chair: Robert Mathieu, National Institute for Science Education
Panel Facilitator: Elaine Seymour, University of Colorado-Boulder
Panel Members: Brian P. Coppola, University of Michigan
Eileen L. Lewis, Canada College
Richard Tapia, Rice University
Commentary: Daryl E. Chubin, National Science Foundation

4:00-4:15 *Break*

4:15-5:30 **B. Small Group Discussions**

Tuesday, February 24

Panel 3: The Role of Evaluation in Institutional and National Policy and Practice

8:00-9:30 **A. Panel Discussion**

Panel Chair: Ann Burgess, National Institute for Science Education
Panel Facilitator: Clifford Adelman, National Institute for Science Educatio
Panel Members: Jack Bristol, University of Texas-El Paso
Manuel Gómez, University of Puerto Rico
Sheri Sheppard, Stanford University
Commentary: Larry E. Suter, National Science Foundation

9:30-9:45 *Break*

9:45-11:00 **B. Small Group Discussions**

Closing Session

11:15-Noon **Reflection and Synthesis**

Chair: Andrew Porter
Cora Marrett, University of Massachusetts–Amherst
John Wiley, University of Wisconsin–Madison

Opening Keynote: Shaping the Future

Luther Williams
Assistant Director
NSF Directorate for Education and Human Resources

I am pleased to join you again. I had the opportunity to make comments at the first two Forums and certainly look forward to this one also being a very productive event. Even though the Institute is only in its third year, it is very important that, after having given the attention initially to the K–12 sector, the Institute moves to extend attention so that it is actually concerned with the entire continuum. Without any debate, many of the issues that you labeled K–12 in your very excellent summary are not K–12; they are at least K-16. I also applaud you for focusing on a very manageable agenda. In an arena as diverse and complex as undergraduate science, mathematics, engineering, and technology education, you did not concentrate on the areas in which there is an abundance of knowledge. Rather you engaged a topic about which I would submit we know very little: success in general, particularly the indicators of success that have been tested in any substantive fashion. What we know in a tested fashion in this arena is almost zero.

I have three points to make. First, the work that you are going to take up here, focusing on the undergraduate sector, maps very well with a report that the Directorate for Education and Human Resources (EHR) commissioned, *Shaping the Future: Strategies for Revitalizing Undergraduate Education* (NSF 98-73). The work was done by the Advisory Committee to the Directorate, led by Mel George. Several years ago we asked this very eclectic group of scientists, engineers, mathematicians, and educators to consider the needs of all undergraduates attending all types of U.S. two- and four-year colleges and universities. To do that, obviously, the study had to be concerned with the needs of majors in science and engineering, with the preparation of K–12 mathematics and science teachers and their continued professional development, with the needs of the individuals who enter the technical workforce, and with the issue of science literacy in general.

The recommendations of the Advisory Committee were that all students have access to supportive, excellent Science, Mathematics, Engineering, and Technology (SMET) undergraduate education and, to the extent possible, that all students learn these subjects with direct experience with the methods and the processes of inquiry. Both recommendations are very challenging. The second one, in particular, requires the aggregate output of the expertise represented by the people gathered here. Certainly, we have not yet established supportive, excellent SMET undergraduate education for all students, and occasionally that education provides little positive attention to processes and methods of inquiry.

The Advisory Committee made a series of recommendations addressed to a variety of sectors, but some of those recommendations were to NSF, and several of those are relevant to this conference. It was recommended that NSF, employing all of its resources, especially those in the research directorates, increase research on the undergraduate educational experience in disciplinary and interdisciplinary contexts, focusing on efforts to integrate research and education. More generically, it recommended that NSF provide support for a research agenda on human learning at the undergraduate level; use the results to evaluate programs, both long term and short term; and develop indicators of success to guide future program development.

My second point is that the agency has attempted to respond to most of those recommendations. We have placed increased emphasis on integrating research and education. Drawing on the resources of EHR, joined with the efforts of the research directorates, the research agenda is being pursued in a variety of arenas. One of the other recommendations has to do with supportive and high quality education for all students. An example far removed from Education and Human Resources (EHR), but extremely important, is the recent decision by the Engineering Directorate to make the transition from the Engineering Coalitions, which you are familiar with, to the daunting effort of systemic reform of undergraduate engineering education. Of course, that's an untenable agenda pursued by engineering alone, because one has to give systemic attention to all of science and mathematics to accomplish reform of engineering.

In EHR and the rest of the directorates, the broad interdisciplinary theme of Knowledge and Distributed Intelligence seeks to integrate the knowledge base for issues from the neurosciences—the human brain, its functions, its patterns of expression—to the important research bearing on cognitive processes, to how one employs learning technologies in artificial as well as human settings, and finally to how students learn in different physical settings, cultural settings, and so on. Integral to that research, it seems to me, is to couple it with what it portends in terms of outcomes. Stated differently, how does one, with the enhanced knowledge base, reconfigure the programs, and then, how does one assess their value in terms of student learning?

We are very concerned with disseminating the excellent products that you are responsible for, and we have used rather traditional methods to do so. But presently we are devoting attention, in collaboration with the Computer and Information Science and Engineering (CISE) Directorate, to actually bring the dissemination agenda to what we now call a SMET Digital Library, kindergarten through probably graduate education. The principal programs in the undergraduate component of NSF—faculty development, course and curriculum development, and instrumentation and laboratory improvement—are increasingly being configured in a continuum. In other words, we are trying to find ways to enhanc the coherence between faculty development needs and course and curriculum development to optimize student learning. a part of the process needed to implement t recommendations of *Shaping the Future* is under way.

We also need continued advice from you in a variety of other sectors. Last fiscal yea we introduced a process of program effectiveness reviews to serve as a precurso to the Government Performance and Resul Act, a congressionally mandated process by which agencies defend their budgets in an outcomes context. We have conducted a variety of these reviews. I want to referenc two of them that we held recently. One was an examination of the course and curriculu portfolio in the Division of Undergraduate Education (DUE). Another was an examination of the research program withi the EHR Directorate. In combining the results of these two reviews we concluded that we need to find robust mechanisms to increasingly advance the knowledge base that undergirds the investment we make. say it another way, while the Institute's w is exceedingly important, there is somethir less than an appropriate representation of research devoted to practice in the undergraduate sector. (I do not imply that there's an abundance of such research ever in the K–12 sector.)

My third point is, What do points one a two suggest for your agenda? I've already indicated that your focus on indicators of success is important and that we know ver little about indicators. A problem-solving design would justify starting with indicato You have to acknowledge not inputs but student learning. What would be definitive effective, durable measures of success? Th

indicators have to be organized into broad rubrics in two domains. One certainly has to do with student performance as a measure, but another one has to be system based. It has to assess the effectiveness of the system responsible for making provisions for quality, supportable SMET education. The two domains interconnect, but one obviously doesn't substitute for the other. In my judgment, that's a major challenge.

My purpose in raising this third point is to bring to your deliberations some emphasis on system thinking. Assessment framed in terms of outputs, that is student learning, does not frame improving undergraduate education. The issue is a question of specificity. Obviously individual institutions, programs, and departments need strategies and mechanisms to report their performance. These reports need to be aggregated in some fashion by which we can assess the SMET enterprise in the nation. Here again I'm emphasizing that student learning and improved undergraduate education do not have the same set of indicators. Whatever mechanisms are used will have to be thoughtfully developed, so that we don't limit the possibilities or constrain the design. What would be reasonable indicators of success from a student learning perspective at the undergraduate level? Which students—majors, mathematics/science teachers, generic matriculants, those seeking science literacy in general? Undergraduate education is a very diverse enterprise. What is the undergraduate analog to the SAT? What would be effective indicators? We have to be very creative, innovative, and in large measure unconstrained by the self-evident to arrive at indicators that actually are going to be useful. And those indicators should not be independent of the fact that the SMET undergraduate enterprise is dynamic. Matching the indicators to an enterprise that does not represent a fixed set is challenging.

You have a variety of very good examples of the leaders of the change represented in the Forum. I urge you to combine considerations of student learning and of improved undergraduate education so that you do not develop indicators for the enterprise of years past. If it is taken up, the combination would allow you to link your findings to the research that has already been done on the K–12 sector. It would be consistent with what I anticipate will be some of the major issues that will be addressed in the science and engineering Graduate Education Forum coming up in June.

I thank you for your continued engagement in this arena, making very clear NSF's continued interest and support. I look forward to the products of your work. Thank you very much.

Panel Discussion Summaries and Commentaries

Panel 1: Assessment of Teaching, Learning, and Curriculum Change in SMET Classrooms[1]

Panel Discussion Summary
Susan B. Millar

Panel Facilitator: *Brock Spencer, Kohnstamm Professor of Chemistry and Associate Dean of the College, Beloit College*
Panel Members: *Diane Ebert-May, Director and Associate Professor, Science & Mathematics Learning Center and Department of Biological Sciences, Northern Arizona University*
David B. Porter, Professor, Department of Behavior Sciences and Leadership, U.S. Air Force Academy
Eric Mazur, Gordon McKay Professor of Allied Physics and Professor of Physics, Harvard University

Abstract

Panelists used a broad definition of course-level assessment, suggesting that assessment practices need to be woven through the fabric of the learning process rather than be conceived as involving only graded evaluation activities. They explained that assessment practices need to be aligned with faculty and institutional goals for student learning, and that good assessment methods help both learners and teachers become more reflective practitioners. They outlined three steps: decide on course goals, communicate these goals to students through both word and example, and use teaching strategies designed to achieve the goals. The panelists favored more invitational classroom assessment practices that challenge students to develop higher-level thinking skills without inducing destructive competition among students. They emphasized the importance of fostering student and faculty buy-in and acknowledged some of the risks posed when faculty adopt new methods. They suggested that faculty interested in using these new methods build supportive networks by linking with like-minded faculty and administrators to achieve a critical mass and by being prepared for the ups and downs inherent in the change process. Finally, they noted that course-level changes in postsecondary Science, Mathematics, Engineering, and Technology (SMET) education need to be aligned with K–12 SMET reform more generally.

Spencer began the discussion by inquiring how well traditional assessment methods inform faculty of what students are learning. The panelists responded by considering multiple-choice exams, stating that this assessment method fails to provide instructors with either (1) adequate information about higher-level abilities, such

[1] Although the title of this session suggests a focus on assessment indicators, the panel discussion actually focused on how to implement reform at the course level. The papers written by the three members of this panel appear in Appendix A and provide classroom-based examples of assessment practices.

as critical thinking skills, or (2) substantive feedback about student learning and their teaching. Moreover, the panelists asserted that multiple-choice exams and similar assessment methods create a "conspiracy of mediocrity" that fosters superficial learning. Such traditional assessment methods imply that the instructor's task is to package and transfer information to students who should dutifully receive, retain, and reproduce it in the final exam.

Need for Alignment

Having agreed among themselves that traditional forms of assessment generally are inadequate for current educational goals, the panelists presented strategies for improving assessment activities. They agreed that the essential first step was for faculty to articulate their learning goals before considering assessment activities. They stressed that these goals should be aligned with both student needs and institutional goals. Once these goals have been articulated, faculty can design assessment strategies that are "aligned" with them. This requires, as Ebert-May explained, that "learning outcomes be clearly defined in operational terms by specifying what criteria for performance you will accept at which points in the course to award a C, and what criteria for an A."

The panelists emphasized that, as faculty articulate new learning goals and develop assessment strategies in alignment with these goals, they must also use teaching strategies that serve these ends. Each panelist discussed from personal experience teaching strategies that are commensurate with particular learning goals.

One strategy that Ebert-May has found productive is to communicate to the students both the goals of the course and the assessment criteria—in the form of a scoring rubric—by which they will be evaluated. In so doing, instructors give students road maps for the paths they are about to travel and share with them the responsibility for achieving the course goals. In her view, it is especially important for faculty who are experimenting with new ways of teaching and learning to use this strategy. Extending Ebert-May's point, Mazur argued that, as well as sharing their evaluation criteria with students, teachers should offer them examples of excellent responses to homework and exam problems and other types of work expected of them.

Mazur described a teaching method he uses to achieve his goal of helping students develop the habit of seeking out information, rather than merely receiving it. To help students learn to synthesize information from resources, he requires them to develop a two-paragraph summary of the reading before lecture. On a spot-check basis, students earn bonus points for their summaries.

David Porter advocated the use of teaching strategies that "walk our talk." He argued that "the scientific method is more likely to be caught than to be taught—our students learn a lot from what we do."

> The heart of the scientific method is how we view mistakes: we honestly endeavor to find evidence contrary to our assumptions. We need to do that as faculty when we work with students in the classroom—and also as reformers, when we work with fellow faculty. We need to create a climate where both our own classroom and lab mistakes, and those o our students, are appreciated for the opportunities for learning that they bring. Viewing our work as teachers in this way involves a paradigm shift. Often we assume that we, the learned, are there to give something to the unlearned, the students. That assumption is in itself a major obstacle real education taking place in the classroom

The panelists also described how they design their assessment strategies so as to mitigate student resistance to new teachin and learning strategies. Each noted that th students most likely to resist new methods are those who learned how to perform at

especially high levels in traditional SMET courses and those who are taking SMET courses as prerequisites for competitive majors. Each panelist offered different assessment strategies.

Porter explained that his assessment strategies are designed to (1) communicate clearly to the students what he values, (2) answer the question, "How is each student good?" by using diagnostic tools that also are intended to engage students in learning, and (3) answer the question, "How good is each student?" by using evaluative tools that inevitably are also anxiety-producing. Thus, his assessment activities convey, encourage, and reward the new kind of learning he seeks to achieve.

Mazur starts his course with a questionnaire that asks students, in part, to describe their course goals. In the next lecture he presents an analysis of their goal statements (contrasted with his own goals) and suggests that the course will work better if the students' and instructor's goals match. A month later, he administers a second questionnaire that includes the question, "If you were teaching this course, what would you do?" Typically, he learns that most students say they would do what he is doing and only describe a few things that bother them. He then addresses these problems in the next lecture. (Details on this method are presented in Mazur's book *Peer Instruction: A User's Manual.)*

As already noted, Ebert-May spends significant time at the beginning of a course explaining her course goals. Throughout the course she requires the students to engage in self-reflection about the degree to which they think they are accomplishing these goals. Self-reflection encourages students to take responsibility for understanding what they need to do and how they need to proceed in order to accomplish the goals. She finds that even resistant preprofessional majors begin to function as reflective learners: "Their attitudes change 180 degrees," constituting a paradigm shift not only for the students, but also for the faculty members whose reform efforts are strongly reinforced by positive changes in their students.

With respect to the idea of fostering reflection as a learning strategy, Clifford Adelman, a participant, described how using an experimental examination called "Academic Competencies in General Education" fostered faculty reflection.[2] A faculty panel developed assessment questions by means of sorting, by cognitive complexity, large numbers of student answers to open-ended questions. This process led the faculty panelists to become more reflective about how students actually think and to realize the need to help students recognize naïve assumptions and use the scientific method to arrive at verifiable assumptions. Subsequently, in a half-dozen workshops, faculty from some 70 institutions used the process, principally for faculty development. Adelman commented that the purpose of assessment at the classroom level is to encourage both students and faculty to engage in analysis processes that help them understand the students' current and evolving assumptions and knowledge. This purpose contrasts with the more traditional purpose of getting all the students to demonstrate that they can perform the same set of predetermined tasks.

Ebert-May concurred with Adelman: she has successfully used a similar sorting process with departmental peers to develop a shared and grounded understanding of what it means to give a grade of C, B, or A. She is finding that those most interested in participating in these sorting processes are the research-oriented faculty. She also observed that, "hallway discussions are becoming livelier and focus less on 'What content did they know?' and more on 'What is the whole breadth of criteria for [what] our students know and do in our courses?'"

[2] This examination was developed in the 1970s and 1980s by Jonathan Warren and was sponsored by the Educational Testing Service.

Addressing Risks to Faculty innovators

The panelists also explored the risks to faculty of shifting to new learning goals, teaching strategies (including assessment strategies), and student outcomes. This topic was introduced by Brock Spencer and sustained by several pointed comments and questions from the audience.

In response to expressions of concern about the risks of innovation to faculty, Mazur described the early stage of implementation as posing the greatest problems. In his experience, initial efforts at innovation often result in lower student grades, which, in turn, may translate into a lower ranking on traditional faculty evaluation forms and into departmental pressure to revert to traditional methods. Thus, faculty reformers face risks created by peer and student disapproval and by peer discouragement when improvements are not immediate. He encouraged faculty to "stick with it," citing the analogy of a tennis coach's efforts to change an athlete's grip: at first, the athlete is awkward and frustrated, but eventually achieves a higher level of performance than would have been possible with the inferior grip she had perfected before the coaching. "Good students" are like the tennis player: while facile when using an inferior method, they will never achieve higher levels unless they rise to the challenge posed by methods that require deeper conceptual understanding. Like the tennis player, faculty too will never achieve higher levels in their teaching unless they stick with it.

Audience comments revealed that many faculty involved in changing their individual courses believe they may be risking or wasting their time unless these courses are part of a sustained or systemic change. The panelists agreed that it is risky for faculty to invest their time in isolated efforts and offered strategies for overcoming this risk. As Mazur put it, "We're in a situation where we have to move a lot of pieces of the puzzle at the same time." All three emphasized that sustained and coordinated change will prevail only if a critical mass of faculty is involved.

Ebert-May encouraged Forum participants to enlist a critical mass of reform-minded faculty through a collegial, invitational approach. Invoking a virus metaphor, she explained that faculty catch the "active learning bug" when they experience for themselves the pleasure of really engaging their students in learning. She cautioned participants against inadvertently poisoning the well by using a competitive approach, that is, by trying to convince colleagues to change by comparing their own (implicitly superior) methods with traditional methods. This strategy can be as destructive for faculty as curve grading is for students.

Mazur observed that it will be difficult to develop a critical mass of faculty unless those who become engaged in reform are able to stay engaged. Faculty will stick with it only if they

- receive support from administrators and funders in recognition of their extra work,
- overcome the lone-reformer syndrome by seeking out like-minded colleagues elsewhere and work with a regional or national network to develop greater visibility and acceptance for the new methods, and
- obtain agreement from their colleagues on "what it is we want our students to learn, how we measure that, and what different measurement instruments that exist now really tell us."

He emphasized that it is destructive to blindly assess new goals with instruments designed to assess old goals.

Need for Systemic Alignment

The panelists concluded their session by zooming out from their own course-based reform efforts and positioning them within a

broad systemic context. They initiated this shift in scope and perspective by noting that development of a critical mass of pro-reform faculty is necessary, but not sufficient for wide-scale success. Success also depends on many other factors, in particular, alignment between higher education and K–12 SMET reform and public understanding of the importance of SMET literacy. With respect to connecting with the K–12 system, Ebert-May emphasized the important role SMET faculty play in preparing the next generation of teachers for the K–12 system: "What kinds of models we are for the next generation of teachers is a piece of the complex puzzle Eric mentioned." With respect to the need for public understanding of the importance of SMET literacy, Mazur showed a video clip indicating that (at least in some quarters) such an understanding is lacking. The clip shows Senate Majority Leader Trent Lott forcefully stating to a group of 400 high school students that taking four years of science and four years of math in high school was "a waste of my time, a waste of the teachers' time, and a waste of space." Reminding the audience of a point made in Porter's Forum paper—that failure to explore, develop, and adopt new and more effective pedagogies poses a great risk to individuals, institutions, and society, Mazur stated, "Unless SMET faculty and K–12 teachers change their teaching approaches, we will continue to produce students—and citizens—who have little appreciation for the importance of science, mathematics, engineering, and technology." His observation closed this panel session.

Commentary

Norman L. Fortenberry
Director, Division of Undergraduate Education
NSF Directorate for Education and Human Resources

We at the National Science Foundation are fully cognizant of the expertise that resides in this room and across the nation in your institutions. After reviewing the papers that were submitted and listening to the comments raised, my question is, "What can be done from an NSF perspective to support the types of innovations you have been describing?" There have been very exciting papers and interesting presentations and lots of exchange and interaction. I want to build on several of the spoken and written comments.

Diane Ebert-May spoke of a need for higher education to join the reform of assessment. Concurring with the remarks made by Luther Williams, I would say we should more broadly conduct educational research to encompass higher education. Most of the educational research currently conducted is focused on K-12. In addition to the well-established community of educational researchers, NSF's EHR Directorate supports a growing number of science, mathematics, and engineering faculty conducting educational research within the context of their disciplines. The people at this table are evidence of this trend. Research on Educational Policy and Practice is a program administered within the Division of Research, Evaluation and Communication (REC) that affects other EHR divisions as well. This program has the potential to link undergraduate faculty with the results of educational research at the undergraduate level.

David Porter made the points about the need for research to inform practice and about our need to predicate questions of student learning on changes in the individual student. A review of EHR-supported research projects by NSF Director Neal Lane earlier this year illustrates NSF's attempts to address this need. EHR showcased three types of research: traditional research within SMET disciplines, research on human

learning (especially in SMET disciplines), and research on applications of technology. Among the projects discussed by EHR were several supported by the DUE that feature teaching and learning within SMET disciplines—for example, the work in physics by Eric Mazur and Lillian McDermott. We're seeing synergy among the disciplines in diffusing educational innovations, not at the level of richness that we would like to see, but it is a beginning. Other projects indicated that support for traditional research within SMET disciplines is informing curriculum development—for example, Art Ellis's work with Materials Chemistry. All of these efforts support the broader goal of integrating research and education. DUE's Course, Curriculum, and Laboratory Development program will support projects that integrate research (including educational research) with curriculum development and implementation.

Eric Mazur talked about using a publication review model for evaluating student work. This model can be related to parallel examination of how faculty can be evaluated for their full range of responsibilities. For example, at Oregon State University, rather than simply relying on publications, faculty have redesigned their tenure and promotion guidelines to evaluate peer-recognized intellectual work appropriately disseminated. This very broad criterion can stand up to tests of rigor and allows appropriate evaluation not only of traditional discipline-based research, but also of pedagogic work; it is applied to the work of administrators as well as the work of faculty. The criterion came about through appropriate use of terms and recognizing what the academic community really values—not publications as publications, but intellectual work recognized by peers, appropriately disseminated.

DUE and REC, both in the EHR Directorate, are planning a very small conference to bring together traditional educational researchers and those in the disciplines to bridge our knowledge base on the how and why of best practices. Such bridge efforts will assist faculty in achieving Porter's goal of creating rich syllabi that offer students options and opportunities to master the material and to acquire individual and group skills.

Ebert-May spoke about students' need to reflect on their own understandings and abilities in providing feedback to faculty. She also made an interesting case for treating all students as prospective K–12 teachers; in a teacher preparation curriculum, you not only teach content to students, but you teach them to reflect on that content and how they would transmit that knowledge. Most of us agree that you really learn something when you teach it, so we encourage all of our students to think about how they would teach what we are teaching them. NSF is encouraging that inculcation of information, treating all of our students as potential teachers.

But Eric Mazur warned us that great change often results in upheaval and rejection. It is hoped that DUE programs are assisting individuals to surmount these challenges. In particular DUE is supporting adaptation and implementation of proven reforms to new local contexts via our new Course, Curriculum, and Laboratory Improvement program and by supporting more localized efforts within the NSF Collaboratives for Excellence in Teacher Preparation. We also hope that the restructured programs will facilitate more coherent approaches to projects that are consistent with addressing all the elements required to achieve high quality teaching and learning outcomes.

Williams indicated a collaboration between the EHR and the CISE Directorates. DUE recently committed $500,000 to working with the multiagency Digital Libraries Initiative to explore educational applications of digital libraries. The Test Beds in Education activity is focused on exploring the feasibility of developing a national resource that supports curricular and instructional innovation, validation, dissemination, and adaptation. It should be a

wonderful platform for conducting research on teaching and learning.

Porter noted that every system is designed to yield the results observed; in DUE, our effort is to be more conscious about the systems that we design in support of achieving enhanced SMET learning by all students. We are devoting increased attention to the means we use to measure outcomes. We are developing more rigorous quantitative and qualitative metrics. Elaine Seymour indicated that there had not been as much evaluation as we would like in earlier NSF efforts. Across our programs, DUE is requiring rigorous evaluation. We hope to soon take the next step and provide metrics up front that principal investigators (PIs) will be encouraged to use. Available metrics should not only strengthen the evaluation done by PIs from a project-based formative and summative perspective, but also allow greater input to an NSF-focused program evaluation across individual projects. One example of "dual use" data is an online Evaluation Resource Library being developed by REC that we hope will be of great use to PIs.

But perhaps more important than measuring outcomes is being more intentional about the inputs, the educational experiments, that we support. As always, we will continue to provide strong support for individual investigators and novel ideas from the field. Such an approach is a defining characteristic of NSF programs. However, congressional requirements for greater accountability and assessment of alignment of programmatic goals with programmatic (not individual project) outcomes requires that we devote heightened attention to where those ideas fit within the broader portfolio we are supporting. Thus, the questions associated with the NSF review criteria remain dominant. However, program officers must also consider how the projects we support inform our organizational effort to provide a rich knowledge base of effective teaching and learning strategies to the SMET community.

Panel 2: Assessment and the Promotion of Change in Departments, Disciplines, and Institutions[3]

Panel Discussion Summary
Susan B. Millar

Panel Facilitator: *Elaine Seymour, NISE College Level One, Director, Ethnography & Evaluation Research, University of Colorado-Boulder*
Panel Members: *Brian P. Coppola, Associate Professor of Chemistry, University of Michigan*
Eileen L. Lewis, Professor of Chemistry, Cañada College
Richard Tapia, Noah Harding Professor of Computational and Applied Mathematics, Rice University

Abstract

The members of Panel Two focused on strategies that they see having the greatest significance for improvement in SMET higher education. They offered three strategies that are useful at all levels and others that are specifically suited to individuals, departments, institutions, or whole systems. The three change strategies that apply at all levels are using assessment data as a driver and key resource for educational improvement; evaluating and improving current evaluation criteria; and having the will to make change as well as knowledge about how to make change. The panelists advocated three improvement strategies for use by individual faculty and faculty groups: focus on efforts that help students through major academic transition periods; get started using any possible path for course improvement; and have a clear destination in sight at the outset. While acknowledging that change strategies used by individuals are important, the panelists also stressed that such strategies are unlikely to prevail unless supported by change at higher organizational levels. To effect change at higher levels, they argued that changes in departmental values and reward structures are the most critical. They also proposed that departments hire faculty trained in discipline-centered pedagogy to change how we foster student learning and advocated that four-year institutions work with two-year institutions to engender improvements in the quality of higher education for the large number of students served by two-year institutions. Just as they emphasized that change strategies used by individuals are insufficient unless supported by change at department and institutional levels, the panelists explained that department- and institution-level change can only be effective as part of a systemic effort that supports it. All three panelists argued that higher education is morally obligated to play its part in a societywide effort to solve the serious national problem posed by science illiteracy. The solution entails a shift to teaching strategies that present science as relevant and exciting—as opposed to distant and dull. Higher education

[3] The papers written by the three members of this panel appear in Appendix A.

leaders at all organizational levels also must work to align their different value systems to develop and enact effective and coordinated strategies.

Seymour set the stage by explaining that this panel would consider the effect of departmental and institutional values on efforts to improve undergraduate SMET education and on theories of change in higher education.[4] She then asked Forum participants to describe the things that their departments value. All but three answers from the audience indicated a departmental focus on research and faculty autonomy. Of the "minority" voices, one cited increased enrollment as an important departmental value, and two described their departments as valuing the needs of students and society. Seymour next asked what factors Forum participants viewed as facilitators of and barriers to change. Members of the audience included among facilitating factors the high value faculty place on student learning, leadership at faculty and administrative levels, regular conversation and collaboration between SMET faculty and faculty in other disciplines, pressure from students who have had more exciting courses, and external money for faculty development. Barriers included the reward system, faculty hierarchy, and lack of money to support the time faculty need to revise curriculum and develop new pedagogical and student learning assessment techniques.

With this audience prethinking exercise as their launch point, and guided by occasional questions from Seymour, the panelists articulated strategies for promoting change in postsecondary SMET education, occasionally also stating the change theories on which these strategies are based. Most of the strategies that they advocated were intended for use at particular organizational levels, individual to systemwide, and a few were intended for use at all levels. This summary begins with the latter and proceeds to the former.

Change Strategies that Apply at All Levels

1. Assessment data should drive reform. The three panelists asserted that at all organizational levels, from the individual to the national, assessment data are an essential resource for and driver of educational improvement. Both Lewis and Tapia saw current efforts to improve college-level SMET education as hampered by the lack of basic assessment data. Faculty need assessment methods that can ascertain what students do and do not know, where their problems lie, and how best they can be helped. They also need assessment data to demonstrate that reform is making a positive difference to student learning.

According to Tapia, the NSF also believes that collection and use of evaluation data are essential factors in successfully institutionalizing improvements. It seeks to act on this belief through the design and implementation of its grant programs. He also observed that NSF has not yet been successful in enforcing policies requiring PIs to make effective use of evaluation. A critical problem in implementing these policies is that SMET faculty do not yet know how to conduct program evaluation and need help in developing student learning assessments appropriate to their course learning objectives. Both Lewis and Tapia emphasized the importance of using third-party evaluation experts when seeking to provide evidence of the efficacy of education reforms,

[4] More explicitly, the panel had been charged to address the following three issues: (1) Learning assessment and reform: What role can the assessment of learning play in undergraduate education reform—in departments, among colleagues, and within the disciplines? (2) Assessment as argument: How do we best leverage change in these spheres? Can assessment data be used to convince colleagues, protect innovators from risk, and build support for educational change? and (3) Effective dissemination: What forms of dissemination work best in encouraging others to try, to support, or (at minimum) not to obstruct classroom innovation?

because outside evaluators provide more objective data.

Daryl Chubin, the panel discussant, responded to Tapia's remarks on behalf of the NSF. The NSF, he explained, encourages the use of external evaluators for education reform projects to help grant recipients think systematically about what they are learning and to give faculty, who are so close to the projects, someone else to help them develop the lessons. Chubin added, however, that evaluators are not likely to be effective at dissemination and that PIs should enlist the help of communication specialists who can devise strategies needed to communicate the lessons learned to diverse audiences, including the department, campus, region, and academic discipline associations.

While strongly endorsing the critical role of assessment data in higher education reform, Coppola cautioned that widespread appreciation of the role that high quality analyses of assessment data can play in improving education will require cultural changes in the educational establishment. Noting that people do not assign value to things they do not understand, he observed that changing this feature of higher education culture will require a critical mass of faculty and administrators who (a) have access to evaluation studies that analyze the interactions among the many critical factors involved in making change and (b) develop an appreciation of the kinds of evidence that assessment and evaluation practices can offer.

2. Evaluating current evaluation criteria. Tapia focused on a second strategy that he believes is critical and applicable to all organizational levels: evaluating the nature and utility of prevailing evaluation instruments. He advocated approaching change "by evaluating whether the commonly accepted evaluation criteria foster or hinder the outcomes we value." While arguing that this strategy should be applied at all levels, he used an institution-level example to illustrate the importance of this approach. He criticized the practice common at exclusive universities of relying heavily on SAT scores, which are a one-dimensional evaluation tool. He explained that a Rice University study found that, although this admissions instrument can predict that students scoring below 850 are unlikely to do well, it cannot predict the likelihood of success for students with scores above a threshold of 1050. On the basis of this study, Rice University developed strategies to compensate for the shortcomings of the standardized admissions exam. First, SAT scores are used only as a "threshold indicator," and other indicators are used to select from among students who pass that threshold. Second, all new students are required to take diagnostic exams. These exams help Rice to place talented students who lack certain skills or knowledge into courses that will give them the best chance to fulfill their potential. Tapia reported that this combination of strategies is working: most Rice students graduate and go on to successful careers in industry and government as well as academia.

3. The will to action. Coppola noted a third change strategy that applies to all organizational levels: knowledge of what works must be combined with the will to change. He likened institutions of higher education to physicians who smoke, do not exercise, and are overweight. They have all the knowledge that they need to establish and maintain health, but lack the will. His point was that "both skill and will are needed."

Strategies Effective for Individual Faculty and Faculty Groups

The panelists advocated three grass-roots strategies in which faculty can engage on their own or in small groups.

1. Focus on student transitions. Tapia promoted a strategy of change that places students at the center of attention. In pursuing this strategy, he has found that faculty can maximize their positive impact on students—particularly minority students—by

selectively focusing their efforts on students who are in transition. By arranging for support groups and providing mentoring during the six- to nine-month high school/college and college/graduate school transition points, faculty can help these students "go perfectly on their way."

2. *Just get started.* Lewis observed that there is no one way to go about reform and advocated a just-get-started approach to change. The important thing is "to take those first few steps," regardless of one's level of knowledge, commitment, or understanding of where the initial steps will lead. Speaking as someone who has gone down this path, she explained that

> you become committed to this journey in one way or another. Once you have been in a classroom where the students are talking about ideas, where there is dynamic interaction, where you see the excitement on their faces, you can't go back to a classroom where they are all sitting there with the glazed look, writing notes that, as someone said this morning, go straight from the instructor's lecture notes to the students' lecture notes with no cognitive processes in between.

Once engaged in this process, Lewis observed that faculty will go on to seek out and even create the resources they need to further develop their new teaching strategies and spread them to colleagues. They seek out like-minded colleagues, investigate the emerging literature on postsecondary teaching and learning, team up with science education colleagues who have expertise in teaching and learning and evaluation, attend and give workshops, and gather and use evaluation data. By drawing on these resources, SMET faculty can avoid going down blind alleys and can improve their effectiveness.

Lewis theorized that the people most likely to start moving down a path that engages them in more interactive approaches to learning are those who, for one reason or other, are on the sidelines or have a different agenda. According to her "fringe theory of change," such instructors tend "to listen more to students . . . and really pay attention." As a result of this different type of interaction with students, they "begin to get a different view of learning" and begin to "do small things differently."

3. *Have a clear goal.* While agreeing with Lewis that many different promising higher education change strategies are evolving, and that any one of them might work well for a given person, Coppola disagreed with her just-get-started approach. He believes that it is critical to start with a clear destination in sight. He concurred with the point made in Panel One that alignment of goals, instructional strategies, and assessment practices is critical to successful education reform, but offered one caveat: the greatest priority should be placed on articulating the goals. Stressing once more the value of assessment data, he observed that, while these data "can tell you exactly where you are, [they] will do you no good unless you have a clear notion of destination."

Departmental and Institutional Strategies

While emphasizing that change strategies used by individual and ad hoc faculty groups are important, the panelists also stressed that grass roots reform efforts ultimately will amount to little unless they are supported by change at higher organizational levels.

1. *Changing the departmental reward structure.* All three panelists strongly agreed with Tapia's assertion that rewards both "drive the whole system" and "are determined at the department level." Tapia supported these statements by noting that college presidents do not affect faculty behavior by exhorting faculty to foster student learning; it is the departments that determine faculty behavior by making the tenure and funding decisions, and departments are not held responsible for realizing campus vision statements. Funding organizations such as the NSF and AAAS

also affect faculty behavior because "funding confers credibility." The bottom line, therefore, is that departmental rewards are critical in effecting a sustained change in teaching practices. He reluctantly advised junior faculty not to endanger tenure through involvement in education reform early in their faculty careers because those who are not there can't help change the system. Tapia reiterated the need to change the departmental reward system to meet the needs of *all* students. He argued that achievement of this goal will require the commitment of one or two members in every department, and that this level of commitment will not be achieved without a change in the reward system.

2. Recognizing the contradictory effects of funding. Coppola elaborated on the impact of funding practices on education change. He observed that, when funds for a change program are provided by an external source and the department fails to assume the program cost in the end, external funding can act as a double-edged sword: even if evaluation data demonstrate that the program is effective, it loses credibility when it loses funding. Coppola observed that, in a grants-driven "funding confers credibility" culture, the take-home message is that "what is credible is getting on to the next funded thing."

3. Recruiting discipline-centered specialists. Coppola observed that departments can use faculty recruitment decisions to foster change. He hypothesized that, just as departments continuously shift the direction of their research interests by hiring faculty trained in collateral disciplines, institutional change in "how we foster student learning will come by treating discipline-centered pedagogy as an emergent area within the disciplines and hiring faculty with expertise in that area."

4. Collaborating across two-year and four-year institutions. Shifting the focus from cross- or multidisciplinary interactions within departments to interactions across types of institutions, Lewis argued that collaborations between four-year and two-year institutions can act as a powerful lever for change in educational practice. She believes such cross-sector interactions are critical because half of all U.S. college graduates begin their college careers in two-year institutions, and because these institutions, although teaching-focused, are inclined to adopt the pedagogy and curriculum used in four-year institutions. She observed that, however promising it may be, this cross-sector change strategy has been seriously hampered by a history of weak communication between four-year institutions and community colleges about reform and assessment goals.

Systemic Strategies

Consistent with their view that the change strategies of individuals and ad hoc faculty groups will founder unless supported by change at department and institutional levels, the panelists argued that change at these levels will be effective only if it is part of an encompassing systemic effort.

1. The necessity of system change. Coppola argued that even if many individuals, departments, and even institutions work continuously to remediate students and faculty who have been "damaged by a flawed system," the underlying problems will persist. He believes that we will achieve science and mathematics literacy for all only by changing the system in which the students and faculty are trained.

2. The offering of science for all. Seymour asked whether the generally elitist and meritocratic traditions of higher education are in conflict with reform efforts motivated in part by the goal of enhancing science literacy. The panelists responded that, although a conflict may be implied by current higher education practices, the mission of higher education certainly is not in conflict with this goal. On the contrary, higher education will fail to achieve its mission unless it plays its part in a societywide effort to achieve this goal. Coppola argued that

taking a meritocratic approach in introductory science courses is actually immoral, in that it "accomplishes a local maximum [by selecting for the very best students] at an incredible price." He believes it is entirely feasible to foster effective learning for all without harming the exceptionally high performers. Lewis stated that higher education can no longer afford to present science in terms of abstractions, but must design introductory college science courses in ways that enable people to understand the impact that science has on their lives. Moreover, she believes that faculty need to make it clear to students that the professoriate is but one of many kinds of exciting science careers that await them. Tapia argued that higher education is morally obligated to foster scientific literacy for all by working to correct the media-fostered nerd image of science that so strongly influences the attitudes of young people, especially those from inner-city areas and from other less educated backgrounds.

3. The social importance of science literacy. Pursuing the role that higher education must play in helping foster K–16 SMET systemic reform, Tapia stressed that higher education leaders at all levels must work with other sectors to address the problems posed by the presence of a permanent underclass in the nation. He observed that conflict in the values and agendas held at different levels of the science establishment is to be expected: what is seen as good at the department, division, institutional, or national level might not be so viewed at other levels. However, while these conflicts may be inevitable, they must not be allowed to prevent us from addressing a problem that "endangers the health of the nation"—the emergence of a permanent underclass whose growth is fueled, in part, by science illiteracy. He concluded the discussion by advocating that people at each level work to understand each others' positions so that they can "align their value systems" and work together to solve this serious national problem.

Commentary[5]

Daryl E. Chubin
Director, Division of Research, Evaluation and Communication
NSF Directorate for Education and Human Resources

I'd like to begin by acknowledging the division of labor: the expertise of the panelists is in classrooms, disciplines, and institutions; they are *content* specialists. I am a *context* specialist concerned with policies and practices. There is a national interest in assessment and a federal role. A grant is a policy tool. Therefore, we might ask, How does agency funding provide incentives for

- opportunities that affect admission decisions (i.e., access to those classrooms and institutions)?
- change (in classroom teaching and learning practices)? and
- participation in the science, engineering, and twenty-first century workforce of the nation?

The panelists have helped us, based on their considerable experience in different academic settings and with a range of students and faculty colleagues, to address the issue of what is valued by an institution of higher education (IHE). An institution's values are reflected in how it decides to admit undergraduate students; how it

[5] Note that Chubin's commentary refers directly to points developed in the panelists' papers as well as their remarks at the Forum.

rewards its faculty for instructional innovation and improvement; how it measures increased student learning; and how it defines "success" of undergraduate preparation for advanced training, workforce entry, disciplinary leadership, and career uncertainty.

Tapia focused on flawed evaluation tools or their overinterpreted use in the admissions process. The traditional barriers, and first-hand experience with their erosion that he described, remind us that applying the wrong filters restricts opportunities—especially for those from groups chronically underrepresented in science but demographically ascendant in the U.S. population.

The "focused prodigy" profile, as he called it, selects on a best-and-brightest model that overvalues the manifestation of early aptitude, denies individual differences in cognitive development, and predicts success—however defined—for but a narrow band of the student population.

This model illustrates the difference between valuing what we measure instead of measuring what we value. Tapia acknowledged the flawed metrics at our disposal, but offered a reasoned interpretation of them, i.e., the use of a threshold instead of attributing quality differences that are hard to defend quantitatively. An SAT score of 1050 tells Rice University about one key dimension of an applicant's ability. Then admissions officers look elsewhere to inform the decision about who will succeed academically.

But what is "success"? What are we trying to predict: first-year GPA, persistence in the major, completion of the baccalaureate in less than six years, capability of doing graduate work, making original contributions to disciplinary knowledge, or less tangible characteristics such as motivation, ability to work in teams, flexibility (all marketable skills), or capacity to learn through the life course?

As Donald M. Stewart, president of the College Board, observed in a recent *Chronicle of Higher Education* opinion piece (1/30/98):

> Dropping the SAT may seem to be an expedient, short-term solution to a long-term problem, but it is also very shortsighted. SAT scores provide a vital piece of information about a student's ability to perform college-level work. A 1993 study established that the SAT is a good predictor of college performance for all ethnic groups, including Hispanic students. . . .
> . . . Although it is crucial to maintain fairness near the end of the educational cycle—in college admission—it is equally important to deal with the shameful unfairness that many children face at the beginning: . . . tracking . . . , poor teaching, and inadequate spending on facilities, books, and other educational resources. Helping students get into college begins not at the college door, but at the schoolhouse door. . . . Race, ethnic background, or family income can still limit students' educational future.

The link to K–12 preparation, made real by college admissions decisions, reflects the uneven playing field that affirmative action was designed to level. Current legal challenges put the nation at risk of losing the strength of diversity at succeeding stages of the education system. Assessment is central to passage through that system.

Coppola's challenge is whether we can realistically expect assessment to become a mechanism of cultural change in IHEs, once students have been admitted. He presented an intriguing bottom-up, faculty-based, department-led argument for change. He admitted that the "demand for accountability has driven the current assessment movement." I sense he would prefer that faculty—those he calls the sole caretakers of acceptable practice—rethink the form, content, and uses of assessment for noble purposes: faculty empowerment, self-improvement, and student learning that goes well beyond a focus on disciplinary knowledge.

He raised the issue of rewards explicitly—teach well but not too well—that accurately depicts the tradeoff between research and teaching. How a tenure-track faculty member spends his/her time, conforming to the expectations of the prevailing department culture, is a dilemma for departments and institutions to resolve.

It's worse than that, he pointed out. Disciplinary knowledge is the context for teaching, not the connections of discipline to other disciplines, issues, and social problems. The license to make such connections comes with teaching the courses for nonmajors. Required courses for the major can insidiously narrow the scope of classroom assessments by, for example, emphasizing certain skills to the exclusion of others. (See his paper for additional examples.)

Coppola's six categories of assessments of student and faculty performance deserve serious consideration, which I urge that you give them. Suffice it to say, there is a need to see testing as an outgrowth of pedagogy—as process and product—and to decide what is meant by and measured as "learning."

If an IHE defines a campus community, then a "public discussion of pedagogy," as he called it, is a minimum requirement—beyond what any faculty or department does. This kind of self-examination requires more reflection than science and engineering faculty tend to engage in. Solving problems "out there," not fixing deficiencies "in here," is their mission.

The devolution of responsibility to the smallest units on campus—departments and programs—makes it difficult for students, parents, and the larger community to understand what the institution stands for, what its commitments are, and how its values will make a difference in the lives of its graduates.

While Coppola found the "understanding of science education" a primary goal, sponsors such as NSF need better assessment information to help demonstrate improved outcomes from investments in faculty development and student achievement. Educators and sponsors alike need evidence of "institutionalization and cultural change."

Lewis indicated that, if NSF's *Shaping the Future* is to warrant undergraduate reform, then evaluations of projects funded to advance this cause are critical for capturing "curricular changes." Note that, in community college settings, resistance to change in classroom practice stems not from a higher value placed on another activity, like research, since these are foremost teaching institutions. Rather, instructional innovations are seen as integral to the institutionwide goal of faculty development.

Buy-in from the faculty can be aided by a clear institutional message that change in teaching and learning is valued. But I wonder whether your presence at this conference makes all of you atypical—more inclined to change and more likely to lead the reform process on your respective campuses.

K–12 teacher professional development has taught us that different strategies are needed with the cynics and the laggards back home. If reform is to take root, then the change agents who are decidedly outnumbered must be fortified by a repertoire of ideas.

One size seldom fits all. The sharing of models that work and innovations that didn't are valuable assets—especially since UCLA Cooperative Institutional Research Program data show that our top high school graduates who enter IHEs as biology or mathematics majors persist in those majors at appallingly low rates from freshman to sophomore year. Overall, women are retained less than men in all science and engineering majors, with the gap in engineering the most glaring.

To summarize . . .

- In *admissions*, I would argue for holding institutions more accountable for wielding selection criteria wisely and affording opportunity that reflects the diversity in the applicant pool.
- In *retention*, the locus of control shifts to department faculty. Their pedagogy and assessment tools are filters for

encouraging and discouraging students. Sorting is inevitable, but using content mastery alone as the yardstick of promise and capability is an assessment system that over-weeds and under-cultivates.

- The *community college* is the repository of our student population's potential strengths and weaknesses—its ethnic composition and math/science preparation. We should think harder, and more seamlessly, about the education continuum as K–14 or K–16, not as K–12 versus higher education.

Finally—a radical suggestion—perhaps the salvation of undergraduate reform will be the out-of-class interventions that provide support, sensitivity, monitoring, and mentoring of students to compensate for those faculty and courses that remain impervious to change.

My hope is that federal dollars can demonstrably foster cultural change and build human capital, especially at institutions that assist faculty in becoming true scholars: customer-oriented assessors, facilitators, and mentors to the next generation of scientists, engineers, and citizens.

Panel 3: The Role of Evaluation in Institutional and National Policy and Practice[6]

Panel Discussion Summary
Susan B. Millar

Panel facilitator: *Clifford Adelman, Senior Research Analyst, Office of Educational Research and Improvement, U.S. Department of Education*

Panel members: *Jack Bristol, Professor Emeritus, Biological Sciences, University of Texas-El Paso*
Sheri D. Sheppard, Associate Professor, Mechanical Engineering—Design Division, and Co-Director of the Stanford Learning Lab, Stanford University
Manuel Gómez, Vice President for Research and Academic Affairs, and Director of the Resource Center for Science and Engineering, University of Puerto Rico

Abstract

In response to the question of how to use assessment and evaluation data to effect change at the institutional level, the members of Panel Three provided a complex, yet coherent, set of answers. First, they advocated that institutional leaders use high quality assessment and evaluation data to assess and then foster institutional "health." In particular, they recommended that institutions make smart use of existing and unobtrusively gathered data, assess student experience across departments, collect selected types of new data, and ensure the high technical quality of their assessment and evaluation instruments and data analysis procedures. Second, they explained that, since institutional change is fostered by certain external pressures, astute leaders can foster change by making these external pressures more visible to campus leaders. The panelists provided examples of how change at their institutions has been fostered because campus leaders have been made aware of the pressures exerted by, for example, outcomes-based accreditation organizations, national funding agencies, national professional organizations, a wider range of criteria used in faculty hiring practices, and competition for student enrollment generated by courses and programs delivered via new information technologies. Last, the panelists stressed the importance of effectively interpreting and communicating evaluation and assessment findings for use by different types of stakeholders.

Clifford Adelman introduced the topics that each panelist addressed and quizzed the audience about higher education information on "the national radar screen." For example, he asked the audience whether they knew the national data on the proportion of students attending more than one postsecondary institution as undergraduates. He inquired whether members of the audience are asking, and know where to turn for answers to, critical questions such as

> Are we providing our students with the knowledge, tools, and skills to succeed in the

[6] The papers written by the three members of this panel appear in Appendix A.

> labor market?...Are we using content analysis techniques to determine how well introductory SMET courses taught in two-year institutions match the expectations of the four-year institutions, and are we using these analysis techniques to leverage change?

Many Forum participants seemed unaware of, but interested in, the questions and data resources that Adelman described.

Manuel Gómez also set the stage by stating that, because the panelists assumed that (1) the Forum participants are here because they want systemic reform, (2) systemic reform depends on cultural transformation at the institutional level, and (3) evaluation and assessment data constitute the feedback mechanisms that drive institutional-level reform, the question they would address should be, How can assessment and evaluation data be used to effect change at the institutional level?

Use High Quality Data to Assess and then Foster Institutional Health

Adelman proposed that, just as one can get "lost in the forest" unless one locates the right "footprints," one can get lost in the wide array of available data about faculty and student activity unless one knows which data provide effective indicators of the parts of an institution that are working well and those that are weak or broken. Developing Adelman's metaphor, Gómez proposed that each institution needs to take a systems engineering approach to assessment:

> System engineers use a research model. They look and finally come to three or four key elements that can be seen to constitute the macro-variables that describe and can help drive the whole system. These variables may be commonplace, like enrollment and graduation data. Once you identify them, you then trace out the implications for the subsystems. This task is difficult, but it can be done. It is the process of identifying the right footprints that guide the path through the forest.

He noted how effectively an integrated approach benefits systemic initiatives.

> A problem with the systemic initiatives is that their evaluation is not systemic . . . . It looks a the little parts, not the whole. [The data are] itemized, not integrated. . . . As for higher education, how much feedback and interactior goes between departments and deans and vice presidents of research? A dysfunctional systen is one where each unit thinks for itself. An effective system is more than the sum of its parts and has systems engineers who look at the institution in a holistic manner, develop a sense of how the parts are interconnected and interdependent, and assess whether the parts are interacting properly.

Gómez provided an example of how information on key variables led to change at the University of Puerto Rico (UPR). The UPR School of Science had been blaming students for failing due to the fact that they were frustrated engineering applicants. Upor receiving data indicating that these same students could thrive in science, they began to examine their own program, rather than blame the students. Gómez captured the essence of this story by noting, "You cannot fix something until you realize that it is broken."

The panelists presented the following specific strategies as elements of an effective systems engineering/assessment approach to transforming an institution.

1. Make smart use of existing and unobtrusively collected data. Both Adelman and Gómez strongly advised making wise use of existing data to assess how various parts of the system are working. Adelman also acknowledged that it often is not easy to make use of existing data. "How many of your campuses," he asked rhetorically, "have student record systems that you can (a) access, (b) understand, and (c) pass on to colleagues in another institution in a form they can understand?" Although he agreed that it is not easy, Gómez demonstrated that it is possible to use existing data effectively. Using existing institutional enrollment and

graduation data, UPR staff learned that, during the period in which curricular change efforts were underway, the number of graduates had increased even though enrollment was decreasing. Their data implied that the graduation rate had increased supporting the conclusion that the university system was operating more efficiently than had been realized.

To address the question of whether quality had decreased as the graduation rate increased, staff turned to data that could provide a proxy for program quality—the percentage of their new baccalaureates who went on to prestigious graduate schools. They found that this percentage had remained constant. Gómez also emphasized the value of using sampling techniques to minimize the cost of analysis without sacrificing validity and reliability. In this regard, he advised the NSF to require grantees to include samples of longitudinal institutional data in their grant reports. Such a requirement, he predicted, would result in "change in the critical unit of change, which is the institution." While agreeing that longitudinal institutional data are of much value, Adelman cautioned that these data should be used with care: they do not provide a complete picture because, for example, they rarely track the academic pathways taken by the large proportion of students who transfer across institutions.

2. Assess the student experience across departments. Sheppard emphasized that, in addition to using longitudinal institutional data, it is important to assess student learning processes and outcomes across departments. She noted that the importance of this type of assessment data has grown due to multi-departmental participation in the production of instructional technology-intensive courses. She explained that

> computer-intensive courses are forcing faculty to collaborate with technical support staff and with faculty from different campus units. When the unit of action is no longer a single instructor or even a single department, new questions arise: Who is the instructor? How do you assess these multiple elements in the learning environment? Who should generate and act upon the assessment data? Who should use the data in deciding how to spend dollars on instructional technology? What units should be included in the computer networks? Should students be required to have computers?

To answer these questions, Stanford University has created the Stanford Learning Laboratory whose mission is to enhance the learning experiences of all Stanford students and to create a model for the judicious use of pedagogically informed learning technology. Stanford Learning Lab staff are certain that in-depth assessment is critical to achievement of this mission. Ad hoc experiments and flirtation with emerging technology will not yield the kind of systematic understanding required for efficient deployment of new technical and behavioral learning models. Sheppard concluded that "a new view of assessment that goes beyond content materials and single departments is emerging."

3. Collect selected types of new data. Gómez explained that change also can be fostered effectively by gathering new types of assessment data, but only when there is evidence that the effort involved in collecting these data is likely to lead departments and institutions to undertake, or at least consider, change. A prime example of this type of data is that collected by the Hestenes "force concept" pretest/posttest instruments. David Hestenes, a physicist at Arizona State University, developed these instruments to measure how well students have learned basic physics concepts, rather than how well they can solve algorithmic problems. Gómez reported that a chemistry reformer at UPR decided to adapt Hestenes' approach for chemistry, challenging her faculty colleagues to use a common test containing both depth-of-understanding and traditional items. They found that students enrolled in a section using a cooperative learning and hands-on

discovery approach averaged 80%, compared to 20% for those in the regular sections, on the depth-of-understanding questions, and that both groups performed about the same on the traditional questions. Faculty are planning to conduct tests two years later to assess which group knows and remembers more key concepts from the course. He indicated that these data are worth collecting because they explore issues that most faculty care about. Gómez argued, "I am pretty confident the ones who went through this process will retain more. I think if we can prove that, we'll have a lot of people on board. This forces everyone to look and compare."

4. Ensure technical quality. Adelman suggested that, unless institutions adhere to principles of survey design, construct validity, reliability, and data analysis in creating evaluation and assessment tools and analyzing the resulting data, they may actually find it counterproductive to expend institutional resources on evaluation and assessment. In a similar vein, Sheppard pointed out the negative effects of a poorly designed faculty evaluation tool used at Stanford. For example, because the instrument does not offer a "does not apply" response option, students rank the quality of the course textbook even for courses that do not use a textbook. This same instrument discourages faculty from using new teaching/learning strategies because it only provides students the option of evaluating lecture-based teaching methods. In addition, because data from the evaluation are provided a month or two after the course and in a form that merely ranks faculty relative to one another, discussion among faculty and proactive use of the data are discouraged. Sheppard concluded that, far from providing information that fosters institutional transformation, this evaluation tool leaves faculty wondering about the value of the data collected and feeling irritated by the time spent collecting and trying to make sense of these data.

Bristol stated that his understanding of the problems noted by Adelman and Sheppard had convinced him that top administrators must invest institutional resources in experts who can provide high quality data on student outcomes. Noting that he has become a "consumer of assessment," he argued that, just as research faculty draw on people with expertise in experimental design and statistics, institutions should draw on people with expertise in evaluation and assessment.

Bring External Pressure to Bear by Participating at the National Level

The panelists' second major point, articulated by Sheppard and strongly supported by Bristol and Gómez, was that national organizations and trends are fostering institutional change by encouraging, rewarding, and/or challenging change-makers at individual, departmental, and institutional levels. The panelists presented several examples (below) of how change can be fostered by making these external pressures more visible to administrative leaders.

1. Outcomes-based accreditation organizations. Bristol asserted that national accrediting bodies, such as the Accreditation Board for Engineering and Technology (ABET), the Southern Association of Colleges and Schools, and the American Association of Collegiate Schools of Business are "helping our campuses move into an assessment mode." The new policies of these national accrediting agencies are allowing—indeed requiring—that each institution make its own decisions about the types of outcomes data that should be gathered and analyzed to guide institutional improvement efforts. Bristol acknowledged that the shift from "bean counting" to outcomes assessment entails a learning process that takes time. He also noted that effort must be spent motivating deans and chairs to effectively use the assessment information. However, he believes this investment is worthwhile: it is

leading institutions to understand students' needs better and thus make more informed policy and budget decisions. Gómez illustrated Bristol's point with examples showing how the EC2000 (referring to ABET's new "Engineering Criteria 2000" outcomes-based assessment) approach "operates as a powerful tool for institutional and national change." He explained that the UPR Alliance for Minority Participation (AMP) program, which for some time had been using an outcomes approach, had not attracted the attention of the Dean of Engineering. With the advent of EC2000, the engineering dean not only began paying attention to AMP, but also began actively promoting it. In addition, the engineering faculty have also begun to examine their courses in terms of the students outcomes.

Sheppard reinforced Bristol's point that the EC2000 approach entails a learning process: "Everyone is working now to figure out what outcomes assessment means—how to establish goals for a course, track students longitudinally." She noted the utility of national workshops, such as those held by Gloria Rogers at the Rose-Hulman Institute of Technology, where faculty work "together with assessment experts to figure out the various models for carrying out this kind of assessment." Affirming a point made by Bristol, she cited as one outcome of such workshops that engineering faculty "are realizing that outcome-based assessment is a long-term, ongoing process, not something you do six months before the accreditors come." At the same time, Sheppard pointed to unresolved difficulties for outcomes-based accrediting agencies: they need to improve the processes for training accreditors, develop better ways to assess the relationship between teaching quality (what is really going on in the classrooms) and outcomes assessment, and examine the impact of teaching evaluations on both faculty development and student learning processes.

2. *Education initiatives of national funding agencies*. The panelists affirmed the point developed by Richard Tapia in Panel Two, that the education initiatives of many federal and private funding agencies are leveraging institutional change largely by two means: faculty incentives, based on the idea that funding confers credibility, and evaluation, based on the idea that change should be data-driven. Gómez highlighted the efforts of NSF in "driving assessment" at those institutions in which it is making major education reform investments. At such institutions as UT-El Paso and UPR, institutional administrators "are working very hard to get data to determine whether or not we are being successful."

3. *Initiatives of national professional organizations*. Sheppard explained that some national organizations are fostering faculty efforts to change the culture of teaching by (a) providing opportunities to share ideas and knowledge and (b) making these faculty efforts visible to administrators. A prime example is the American Association for Higher Education (AAHE), which is seeking to help faculty in research institutions across the nation shift "from approaching teaching as an experience that an isolated professor has with students to a collaborative activity among teaching colleagues." She explained,

> A handful of people are really hungry to talk about teaching, but there isn't a convenient format where we can talk about our latest experiment or education grant, or about student learning processes: How do you motivate students? What are more effective ways to foster learning? I think one of the important factors in the AAHE Peer Assessment project (called "From Idea to Prototype: The Peer Review of Teaching") is that the creators of the project recognized that some of these movements need national visibility in order to get buy-in from top administrators. The Peer Assessment project required commitment from the provosts of all twelve universities in the project. The provosts had to commit some time and some dollars to promote this activity. I know that my institution is very conscious of what other schools are doing. The national visibility is an important element.

4. New criteria used in faculty hiring practices. Bristol pointed to another national trend that, while not directly involving assessment practices, nonetheless is pressuring campus leaders to initiate and support change in graduate training practices. He stated,

> I see almost every candidate we bring to UT-El Paso—and they are coming from major institutions. The portfolios they are presenting are different from those I saw eight or nine years ago. They have impeccable degrees in mathematics or physics, but they also have something else. They've experimented with learning strategies. They've worked in summer institutes with minority high school kids in math. This is a very positive sign and it's happening across the country.

He observed that institutions that provide their graduate students with professional development in teaching give them a competitive edge in the job market, particularly at the 4-year comprehensive institutions.

5. Competition for student enrollment generated by new courses and programs delivered via new information technologies. Gómez noted that a new arena of institutional competition for student enrollment has emerged as increasing numbers of institutions offer Web and other electronically based courses and programs. This development is exerting new pressures within institutions: faculty and administrators are realizing that they can no longer afford to operate as "a fairly dysfunctional system where each unit thinks for itself, and you manage to get along." To compete effectively for enrollment in electronic courses, SMET faculty and assessment/evaluation experts need to collaborate. As Gómez explained, "Content and assessment experts need to listen very carefully to each other and look for a design that makes sense. We need teams of experts who work to make a coherent whole. This is a new research frontier."

Tailor Evaluation Information to Your Intended Audiences

Throughout the session, all three panelists developed the panel's third major theme—that effective interpretation and communication of evaluation data are critical. Gómez focused sharply on this issue

> We fail many times by merely giving information, not translating, not analyzing it in terms of factors and concepts that the intended audience cares about. For example, there is the language the professor will understand and the language that the federal system will understand. This constitutes a communication problem. [These days we not only] have to be good at measuring, we also have to be great at interpreting, and outstanding at communicating. I think the la two are the weakest links in our evaluation systems.

Gómez proposed that, to bring about effective institutionwide use of assessment data, one must approach each stakeholder group in terms of its values and needs. As illustration, he described how he encourages faculty to understand and use assessment in a way that leads to education reform. First he appeals to their values and needs: professors are "in the business" because they want their students to learn, but they are quick to dismiss the value of assessment dat on how well students learn for students othe than their own. (He made reference to how findings of a Carnegie-Mellon study showing how little students retain two years after a course did not motivate faculty change.) Faculty must be helped to administer meaningful assessment instruments—like concept based physics and chemistry tests—to their own students. Only when confronted with data on their own students, he explained, will faculty buy into the conclusions and start to change their departments.

Gómez takes a different tack in getting administrators to understand the value of

assessment data. As Vice President for Research and Academic Affairs of the UPR System, he challenges deans and chancellors with such questions as, Is it efficient to have students, on the average, take a course three times to pass it? He then draws on analysis of institutional enrollment data that show gatekeeper courses that students enroll in up to three times in order to earn a satisfactory grade. By presenting assessment data in this manner, he helps chancellors see how such data can help them find ways to invest instructional resources more efficiently.

Gómez then explained that, without effective ways of communicating assessment findings to administrators, change efforts will lapse back to the status quo because "the individual professor who struggles to make a change always becomes a victim of the system's immunological reaction. The lone reformer or a small group trying to make a change is treated like the human body treats a foreign object—you resist it and isolate it." To get the whole system to accept these "foreign objects," it is necessary to locate the macro-variables that efficiently describe the whole system (to pursue the analogy, heart rate and blood pressure indicators), and provide top administrators information that clearly shows the relationships between the change efforts and these system indicators. If these data show how the change efforts benefit the system, administrators can use the data to reframe these efforts as factors that enhance, rather than threaten, the well-being of the system. He therefore urged reformers to structure assessment and evaluation information intended for administrators in ways that clearly communicate how educational change is affecting the system. This communication strategy is powerful because it helps administrators understand when and how they need to redefine system health, to adjust their strategic planning, and redistribute resources in order to foster a healthier system.

In summary, Gómez argued that while change makers at the classroom and department level are essential, isolated individual efforts ultimately will be rejected by the institution if institutional leaders do not understand the cumulative value of their efforts. It is through evaluation and assessment data on key system indicators, and pressures exerted by national agencies and trends, that institutional leaders learn about and then find ways to institutionalize successful reform efforts.

Commentary

Larry E. Suter
Deputy Director, Division of Research, Evaluation and Communication
NSF Directorate for Education and Human Resources

Thank you, panel, for this discussion. I want to thank NISE for organizing and having this conversation. It is a conversation I have been wanting to listen to for a very long time. I am going to try to say where I think we are and where I think we are going. Then I hope to have some more discussion with you.

This Forum asks some very good but difficult questions. This particular panel was to talk about the role of evaluation in institutional and national policy and practice. I found it very difficult to always distinguish the national role from the institutional role. But I think that the organizers clearly understood that, for some people, the change has to occur at the institutional level. However, there is national leadership, which we've all recognized that the National Science Foundation has been providing in the programs that it organizes. In measurement, however, there is not the same kind of

leadership in the higher education area as there is in elementary and secondary education. I want to use a couple of familiar examples to show how I see the policies of elementary and secondary education being organized over the last fifteen or twenty years.

Twenty years ago the issue of student achievement was not primary in educators' minds. It was, of course, the business of education, but it wasn't discussed nationally until the National Assessment of Educational Progress was created and international studies, in comparisons of U.S. students with students in other countries, showed what level of achievement students might acquire. Those studies helped galvanize the education community into thinking and understanding that student achievement is not an inmutable force, but something that can be affected by schools and methods of instruction and school organization. Therefore, policies were created to help think about and experiment with ways of organizing the education institutions. I think that the systemic reform efforts represent an interesting approach, because the ideas of systemic reform come from people in the field. What the people in Washington said is, "Here are the goals, here are the places we want to go. Now you show us how we might get there."

When I came to the National Science Foundation in 1990, I started organizing an indicator report of elementary, secondary, and undergraduate education. Data and information about the undergraduate sector just weren't there. There were lots of statistics, but they didn't provide facts for the leaders at NSF. I could not report how much our students were learning, how the pipeline was working, where the holes in the pipeline were, or whether there was a pipeline, or whether there was a better analogy. The thing that was missing was the organization of information around the subject areas, around content. I think that is what this conference is helping to begin a conversation about. You people are aware and involved with the content of education. If we can get you engaged in discussions about measurement and content, we can make some forward progress.

As a program officer in the Indicators program for a number of years, I tried to encourage studies and proposals from people who might take on the issues of this conference. I received very few fundable proposals. One problem was that proposals came from administrators who were familiar with administering the institution. Their proposals might have facts about school finance—paying students to get to school, looking at participation rates—but these are boring institutional measures. They don't tell me anything about the learning process in the classroom for the effect of certain types of instructional practices, the effect of technology, or the goals of the curriculum for that institution. Kenneth Travers is looking at how measurement might be used as a feedback mechanism to help departments galvanize themselves. That project, and a few others that have been funded recently, might lead to the kind of indicators we are talking about at this conference.

What do I think is next and what is NSF's role? I hope that the Forum organizers will try to summarize the *content* of this discussion, not try to find the issues and report different levels of indicators that are needed. This conference has not done all that we had hoped. It has not laid out a set of indicators. But I don't think anybody can say that is the whole story, and I don't think all of us have made that our goal here. The conference organizers have to say, Where is the forest? How do we find the footprints that are there for the different levels of end users? Footprints for people who are trying to design programs at the national level are different from those for people who are trying to carry out reform at the institutional level and at the department and classroom level. There are probably different kinds of indicators, and we need to recognize that and make some proposals and look at some numbers and hear some further discussion and debate.

At this Forum and in this panel discussion, I also heard a lot of discussion about the need for evaluators who understand evaluation, the need for people who can take a problem and turn it into measurement. In our evaluation program at NSF, we had difficulty finding people to do national evaluations and finding knowledgeable people to do evaluations with local projects. There aren't a lot of people who are willing to tackle the measurement problem and communicate with the subject matter people who know what needs to be evaluated. The biggest problem in evaluation is communicating with the audiences that need to know how to make change, the leaders of the nation or the institution. If the evaluation doesn't make sense to them, it is going to sit on shelves. And if it doesn't make sense to the professional societies that are going to be carrying out the change afterwards, it won't be used. So we need people who can think about the evaluation process and the measurement process, who also understand the subject matter and classroom instruction. Now how do you get that combination? I'd like to see it in one person, but I don't think that is really possible. It is probably necessary to have teams working together, and we need to build those teams. They will probably have to be built at the institutional level, and we need to encourage that. A lot of our program evaluation efforts at NSF are encouraging team building and are trying to create cross-disciplinary communication within the institution.

Future conferences need to be held in which we actually share some true evaluation experiences, some evaluation reports, some real stuff. I want to see some of the results of the efforts for accreditation that Jack Bristol was talking about. That's the new major national movement that needs to be evaluated. We need to have some measures of what the process is and where it is working and where it isn't working. We need measures because they help communicate the process to broader audiences so we can share experiences. We need to gather some consensus about what some of the critical elements are, and then we need to pick up and remeasure from there. One thing that I have learned in my twenty-eight years of trying to develop measures is that the process changes. The processes of evaluation and measurement are not as simple as trying to understand a natural phenomenon that is sitting out there and getting a matrix that absolutely captures it. They are dynamic processes that are partly political and partly a matter of communicating with different audiences—and those audiences are changing. So the content of the evaluation has to be continually evaluated itself and then reflected back on the process. I don't know what kind of national process there is for doing that. NSF is certainly interested in trying to support efforts for improving the evaluation of undergraduate education. I think that our reform efforts are going to drive those evaluation efforts. Thank you.

Synthesis of Participant Think Piece Essays: Voices from the Field—Assessment, Change, and Systemic Reform

Sarah A. Mason, Ramona L. Gunter, Susan B. Millar, and Elaine Seymour

1. Introduction

At the beginning of the small group breakout sessions following each panel discussion, Forum participants were asked to spend 15 minutes writing a brief response to the following question: *Given your experience and knowledge of evaluation and assessment in postsecondary SMET education, what thoughts do you have in light of the preceeding panel?* They were also told their think pieces would be used as a key resource in the production of the Forum Proceedings. Overall, 236 participants wrote think pieces (236 completed think pieces for Panel 1, 192 for Panel 2, and 163 for Panel 3.)

The purpose of this writing exercise was to document and further the conversation fostered during the Forum. While the points made in participants' think piece essays were in part stimulated by the Forum discussions, they also expressed the participants' personal views and experiences. The essays, therefore, constitute a semistructured dataset on the views and felt needs of a fairly broad sample of members of the postsecondary SMET education community who have a strong interest in education reform. From these data, we have produced a synthesis that profiles participants' responses to the Forum topics. The document provides representative quotes from participants[1], and accords more space to those issues most frequently raised.

All three panel discussions focused on the central theme of assessment and evaluation—in the classroom and as related to impact on policy and practice and the promotion of change. Key themes raised by the panelists included higher education culture, support and reward systems, higher education systemic reform, and alignment. Forum participants elaborated on the complexity of these issues in the think piece essays and in subsequent discussions. For example, alignment of assessment to classroom, department, and institutional goals was a predominant theme throughout the think piece essays. Many wrote about the need for developing faculty awareness and understanding of assessment strategies as well as the need for institutional understanding and support for expanding the role of assessment. Assessment was sometimes referred to as a change driver, but was more frequently described as being part of a process and as an important part of curriculum reform and systemic change.

This Forum attracted individuals who, taken together, have considerable experience and expertise in providing high-quality SMET education to postsecondary students. Yet, most of those who wrote think piece essays communicated the belief that this community is in many ways just at the beginning stages of SMET education reform and innovation. In particular, they implied that the community still has much to learn about assessment and evaluation. They noted that it is a good thing that the conversation has begun to bring these issues to the fore and gave credit to individuals who are already trying new innovations in the classroom. Moreover, they appreciated the national dialog spurred by NSF and at conferences such as the NISE Forum.

Others, however, were less confident that the SMET community is on the brink of change. Some advocated the need to re-evaluate teaching, learning, and assessment to determine what works *before* promoting

[1] We identify the writers of quotes only in those cases where written permission to do so was given.

wholesale change and innovation. In writing of the need for developing ways to scale-up, disseminate, and institutionalize alternative assessment and other reform strategies, some of these individuals wondered whether the barriers to such change could be overcome. They noted the need to proceed cautiously and to more carefully define change, promote collaboration, and create a common dialog.

Although many stated that they were fairly new to the concept of assessment and its role in the systemic reform process, their opinions on assessment nonetheless were framed within the larger context of change, systemic reform, and innovation. They wrote primarily about changing the higher education system—both from the top down and from the bottom up. They portrayed the higher education system as interconnected components comprised of classrooms, departments, two- and four-year colleges, and research institutions. Many identified key internal components of this infrastructure, noting the importance of pressure points for change and articulating barriers to and enablers of change. When writing about the change process, they noted the significance of key stakeholder groups, including students, faculty, junior faculty, graduate students, adjuncts, administrators, and evaluators. Many comments expressed the desire to extend the conversation across the science and education communities and across educational institutions. In particular, respondents mentioned the need to include science and math education experts as well as K–12 and community college educators more fully in the assessment and change discussion. Embedded in their frequent requests for examples, models, "how to" specifics, and an expanded dialog was a call for additional opportunities to learn more from assessment and evaluation experts. Participants articulated the need for such mechanisms as communication, collaboration, and dissemination, emphasizing the need for networking and leadership. In addition, they wrote about the influence external groups have on the change process, naming such entities as professional societies, the NSF, and accrediting bodies.

This think piece synthesis is presented in two sections. The first section summarizes the variety of constructive themes regarding assessment and SMET education that emerged from the think piece writing. The second section features the writers' recommendations for building and sustaining change in SMET education.

2. Constructive Feedback: Issues, Views, and Needs

A constructive set of themes emerged from the Forum think piece essays. These themes were expressed in writers' statements of needs and problems encountered in developing and using assessment, enhancing innovative education, and implementing systemic reform. Many writers chose to illustrate a point or issue, or explain a problem, with stories of their own experiences in the classroom or department. Others chose to provide insights into what they thought was missing from the conversation but critical to the larger discussion of assessment. Still another set of responses addressed what was needed, providing an extensive list for Forum participants to consider. Below we present these comments, ideas, and views.

2.1 Defining Change: Purposes, Values, and Process

Many Forum participants wrote of the need for an established rationale, clearly stated purposes, and well-defined values for education reform and innovation in SMET areas. Others mentioned the importance of establishing clarity in the conversation about change, citing the need for definitions, a common dialog, and clear language.

Establishing the Purposes for Change and Assessment

Some Forum participants wrote that the first step needed in making change in assessment practices is to establish a common understanding of the purposes for change overall and for changing assessment practices in particular. A few also believed it is important to establish a shared understanding of what the terms *assessment* and *evaluation* mean.

Susan Ganter and another writer exemplified those who expressed the need for a "common vision":

> One of the most valuable points made in the panel is that goals must be developed prior to assessment. Too often, especially in higher ed., we find ourselves jumping into curricular innovations and then trying to assess student learning and other outcomes as impacted by those changes without ever deciding on the goals of the changes. This problem is relevant to all levels of assessment, from individual student to institutional to systemic. (Ganter)

> Change is a process—not an event—which begins with one's personal vision. This vision is the essence of change. A personal vision may be shared with others (becomes a shared vision). This shared vision is a prerequisite for any anticipated systemic change within an organizational unit.

A few participants urged the SMET community to first determine the purposes for change by first learning more about what works in the "traditional" classroom and what does not work. These writers asked for a similar consideration of new assessment approaches, emphasizing that it is important to balance traditional and new approaches. As one writer put it, "We need to determine if change is required, how much, and for what purposes." Bill Clune also struck the theme that caution is important:

> As with the first session, I found that the discussion of change and resistance to change assumed the conclusion about what kind and degree of change is desirable and possible. . . . Establishing the credibility of the "new" should be seen as an exercise in leadership, institution building, and achieving some degree of buy-in and critical mass. Recall that "old" values are also values—that's how they come to be measured. In other words, we're talking about careful change.

Similar to the call to begin the change process by establishing the purposes was the request several writers made for articulating the purposes for assessment in particular.

> During the panel discussion I was trying to sort out the purposes for assessment at the postsecondary classroom level. . . . Is it to motivate students by making them think about their own personal academic goals? To provide the instructor with information about the course [related] to the particular needs of that class? One of the most important first steps in designing an assessment is to identify the purpose(s) it is intended to serve.

> Is the goal of postsecondary education to facilitate learning in such a way as to produce people who can problem-solve "on the job," or to simply perpetuate higher education? Before appropriate evaluation methods can be implemented, the overall goal of not just the course, but the curriculum itself, must be established.

A few writers related the need to define the purposes of assessment to the need to use a common language in doing so. As one participant explained, "If our goals include thinking critically and understanding the scientific process, we must first clarify what these terms mean for us—they are used [by different people] to describe quite different processes."

Writers also noted that many Forum participants were confused about the distinction between assessment and evaluation, noting that they seemed to be used interchangeably. For example, one participant wrote, "Is there a difference between assessment and evaluation?"

Writers also suggested that clarity in defining the purposes of each would serve to enhance their effective use within the classroom and at the institutional level.

While some writers were confused about the differences between evaluation and assessment, others, such as Barbara Cambridge, referred to the need for using both processes:

> It is important to make the distinction that Cliff Adelman did between assessment that provides information for improvement, and evaluation that uses assessment information and other information to rate and judge. When we make that distinction, we can value everything from the classroom use of assessment for change to the institutional use of evaluation for change—instead of prioritizing one over the other. Systemic change needs to proceed on all levels simultaneously.

Defining Values

Other writers expressed a need to more fully define what is valued in student learning, urging faculty, departments, and institutions to consider: What is it that we want students to learn? Who are we teaching? What do we value? For example, Stephanie Pfirman asked,

> What do we define as long-term success in our students' educational experience: that they are motivated, life-long learners? that they achieve prominence in their careers? or that they achieve an overall balance in their lives that makes them happy and content? At different stages in my life and career I've had different measures of my own "success." What should we be looking to achieve as a national goal?

Eileen L. Lewis and another respondent wanted clearly defined values:

> Our models for assessment and evaluation reflect our models for learning. We must begin by deciding what knowledge and skills we want students to have and then design instruments that assess that knowledge and those skills. We must assess what we value, not what is easy to assess. (Lewis)

> There needs to be important discussion and clarification at institutions about goals, missions, and values. How do the institutions and individuals within the institutions understand the goal to educate the elite (15%) and the goal to educate the [larger] populace?

A few writers believe that the process of discussing assessment criteria can help the members of a department identify and agree on their values and goals. For example, Hubert Dyasi and two other participants wrote:

> I agree [with the members of Panel 1] that assessment serves to clarify goals and the relationship of instructional activities for the students, professor, and colleagues. (Dyasi)

> I loved [Clifford Adelman's] idea of having many faculty sort piles of written work to identify their criteria for "good" learning. I am going to suggest that our multisection introductory courses try this to see if the different faculty involved can reach some agreement on what they value.

> To me, the most useful concept is the realization that assessment must be regarded in some kind of holistic manner and should really be viewed as part of the learning process. We must first decide what goals are feasible for a curriculum and then how the individual courses should contribute to that general goal. Ideally, assessment will contribute to [formulating] a general goal.

2.2 Alignment

Many participants wrote about the importance of a central theme developed by the first panel—that goals for student learning outcomes, teaching and learning strategies, and assessments must be aligned. Think piece writers emphasized that (a) assessment plays a key role in helping faculty and administrators achieve alignment among goals, strategies, and outcomes, and (b) that alignment is critical at all levels within higher education. Depending on

whether they were considering the course, department, or system (institutional and national) level, the participants focused on different issues pertaining to alignment. At the course level, they emphasized how changes in any one component (goals, teaching strategies, or assessments) have implications for the other two components. At the department level, they proposed that alignment can lead to a more coherent relationship among student learning goals, teaching strategies, and outcomes for individual courses and the discipline major(s). With regard to change at systemic levels, they emphasized how change may not be successful or lasting unless relationships among components are understood and aligned.

Alignment at the Classroom Level

Those think piece writers who focused on alignment at the classroom level emphasized that course improvement efforts will be ineffective unless faculty work to align: goals for student learning outcomes; the teaching and learning strategies used to achieve these outcomes; and the assessments used to foster and/or measure the achievement of these outcomes. This general point was made by Ann Igoe, Paul Kuerbis, Hubert M. Dyasi, and Judith Pelchat, among others.

> Goals for instruction and goals for learning must be consistent, and teaching/learning opportunities need to match the goals very closely. (Igoe)

> Assessments and instruction should be closely aligned—as should the content of the course. Our assessments should communicate our expectations for student learning, our view of the important ideas of the content area we are teaching, and how those ideas are interconnected. (Kuerbis)

> It is very interesting that Prof. Gomez thinks assessment drives learning—students learn and study what they think/know will be assessed. Gomez's statement also means the instructor's assessment should be aligned with instruction. (Dyasi)

> Assessment methods must be linked to, and appropriate to, instruction and curriculum. Otherwise, they will make little sense to students and will fail to inform their learning. (Pelchat)

Another subtheme articulated by those who wrote about course level alignment was that efforts to achieve alignment may be ineffective unless undertaken in a sequential way, with goal definition first.

> The need to be clear and explicit about the learning goals for students before attempting to devise ways to assess student progress toward those goals is paramount. (Brock Spencer)

> I think goal setting for each course in the context of the curriculum is a vital first step (Eric Mazur's point). More than content learning as traditionally defined, I believe this goal setting should include: conceptual content to be mastered (including a definition of what that mastery means); identification of skills needed to understand topics and communicate that understanding to others; identification of broader understandings such as scientific literacy (as defined by Arnold Arons); critical thinking; and enhanced appreciation of science. Only then can assessment measures be developed.

> The nature of assessment in SMET depends upon what professors are attempting to achieve in the classroom. Assessment for program improvement is much different if you are attempting to develop skills and methodological understanding rather than merely [impart] facts. The issue of "content" versus "understanding" has to be addressed before the nature of assessment will be clear. What needs to be measured must be known before we can know how to measure, evaluate, or assess.

A few writers went on to suggest that change may even be harmful unless goals are defined first. For example, one person wrote

that, if an instructor does not begin with "the correct notions of what indicators highly correlate with success, one could be trying to drive the incorrect results."

A third subtheme on the course alignment point was that it is important that *students* understand course goals and how the course's teaching/learning strategies and assessments are related to these goals. For example, Joshua Gutwill explained, "Assessments really have to reflect [faculty instructors'] goals—even in the eyes of the student. . . . We need 'transparent' assessments, in which students can see the purpose, goals, and criteria in the assessment." Likewise, another writer stated that she was "particularly interested in how to clarify for students our goals and to align student/faculty expectations more closely." Roger Nanes, Lillian Cassel, and Nancy Romance emphasized in particular the value of considering student expectations in efforts to create courses that align goals, strategies, and assessments.

> Just as we focus on how we need to modify and reevaluate whether there is a match between our expectations and our assessments, we should also devote more time evaluating the match (or mismatch) between student expectations and content assessment. (Nanes)

> Students and faculty must spend time in considering the goals of a given course, project or other educational experience. Without understanding our goals, we cannot construct meaningful measurements of success. If students and faculty have different understandings of goals, students may make substantial progress toward their own goals—while failing to meet faculty expectations. (Cassel)

> What strikes me is the need to be very clear and explicit about the learning goals of any [curriculum]—and within this context the goals must be conceptually organized, sequenced, and articulated. Additionally, these goals (learning outcomes) must be clearly evident to the students. If this is the case, then obviously, the use of new, varied and multiple formats of assessment is appropriate. (Romance)

Kevin Aylesworth, Tony Jacob, and other Forum participants argued that faculty should not only help students understand course goals and their relationship to course assessments, but also design their assessment strategies so that they involve students, thereby more effectively fostering student learning.

> Assessment must be carefully matched with both the desired learning outcome for the students and the course concepts and content. Ideally, they should allow the students as well as the teachers to understand where more work needs to be done. (Kevin Aylesworth)

> To provide ownership of the course by the students (which promotes active learning), either the instructor's goals should be flexible to change with the student input, or the instructor could have some areas where the goals are vague and let the students guide the direction of the course. (Tony Jacob)

> Content, assessment, and teaching strategies must go hand-in-hand. Students need to know up-front what the goals are and how they will be assessed. It would be even better if students help to develop the criteria. Assessment must be embedded into the structure. This would help students become more responsible for their own learning. The assessment strategies should encompass ways for students to assess each other (peer review) and ways for the teacher to assess what students know.

Alignment Within and Between Departments

Writers who focused on alignment issues also noted that it is important that the goals of a course be aligned with those of other courses (within and across departments) and that related courses should incorporate common assessment and evaluation strategies. For example, one undergraduate administrator noted the importance of aligning pedagogical methods within multisection courses:

> My concerns as an administrator are three-fold. The first is to encourage my departments and faculty to begin to consider the implementation of new (or different) methods of curriculum at the general education and advanced (major) level. Realizing that this normally occurs at a course-by-course level, the next step would be to ensure that multisection/multi-instructor courses employ the same (or similar) evaluation and assessment methods.

Marlene Moore and Cora Marrett explained that student outcomes—skills and content knowledge—should be articulated within and between departments.

> It is critical to remember that none of us work in isolation. If goals for our introductory classes do not fit the expectation of colleagues whose courses build on our course, the assessment by colleagues of our course will be negative. (Moore)

> Goals should be set and articulated for any SMET course. Although course-based goals must exist, goals and assessment should [emerge from] communitywide discussion of intended outcomes for SMET education and articulation of assessment strategies across levels and the curriculum. (Marrett)

A few expressed concerns about whether the implementation of innovative course objectives negatively affects student command of course content needed for subsequent courses. This issue was of particular concern to two faculty:

> I am skeptical when the [new course goals] carry with them changes in course content. Are we certain that the resulting course content doesn't lower the standards of the course, affect the smooth transition into subsequent courses, or [hinder] achievement of the type of scientist/mathematician we desire?

> Nagging Question (probably comes from the background of having students who must possess a certain knowledge base to enter a profession): Where and how does curriculum content (amount, type, level) fit in? I think we should not lose sight of this as we begin to focus on higher level skills.

Alignment and Assessment as Drivers of Systemic Reform

Writers also made it clear that, if change is to be viable at a systemic level, it is important that, at all levels within the system, values and goals are explicitly articulated and consciously aligned with assessments and other teaching practices—from the planning stage onward. Assessment within an aligned system can thus function as a "driver" of systemic reform. For example, one person wrote,

> Assessment practices are one of the keys to change. Change the assessment practice and you may change the teaching/learning environment. However, much thought needs to go into what should be assessed. Those individuals who are part of departments, disciplines, and institutions need to consider very seriously what the intended outcomes are to be.

Another writer expressed an understanding of the need for systemic alignment by noting the problems encountered when systems are *not* aligned:

> [That new teaching methods] are being rejected seems to lead to a conclusion that, unless these new methods are more entirely adopted into all aspects of the educational experience, they may be judged in the future as failing. . . . [They] will be judged as being failures because they have been tested in an environment which as a whole does not value a democratic educational model, but rather values a meritocratic educational model.

In short, as Lloyd Douglas wrote, "Institutional buy-in is required at all levels (faculty, departments, the institution) in order for any assessment efforts to be effective."

To achieve this type of buy-in, Terry Millar and another participant argued that diverse stakeholders must be prepared to participate in a lengthy interactive process.

> The alignment of goals and assessment is a critical issue. And it is important to "unpack" the issues implicit in the term "goals" in order to generate departmental, disciplinary, and institutional interest and buy-in. In other words, different stakeholders might have different views on what constitutes acceptable goals, even within one group (say, the faculty). This interest and buy-in will not be a simple, single process, but will come over time as new assessment and evaluation techniques reveal possible goals, and new goals suggest different assessment and evaluation approaches. (Millar)

> Understanding what the values and goals for institutions are and how assessments are aligned with these goals is not a linear process, but rather an interactive process [in which different levels] inform each other, with eventual convergence.

Respondents elaborated on the various roles that assessment and evaluation might play in driving systemic reform at institutional and national levels, noting that assessment and evaluation each can play important roles in the transformation of science and math education. For example, Flora McMartin wrote:

> Assessment must have an organizational role as well as a classroom role. It is not enough to know what works in one classroom—this information must be used systemically.

Another participant wondered whether assessments that are aligned with the new "student-directed" teaching approaches could provide national standards:

> What type of assessment/evaluation instruments need to be created to measure learning in the new student-directed (discovery/inquiry-based, learning-centered vs. teacher-centered) paradigm? Can these instruments be standardized to provide some level of national assessment?

Robert Mathieu explained the important role assessment plays in systemic action by elaborating on Eric Mazur's point that students who are facile when performing lower level cognitive skills will never achieve higher levels unless challenged by teaching methods that require deeper conceptual understanding. He wrote:

> Panel 1 clearly argued that [if students' conceptions of what learning entails are in need of change, then] criteria and standards of classroom assessment are also in need of change. (As Mazur said, his students weren't really as good as his previous exams were suggesting.) However, if the rules change, how do we convince our colleagues and our students and the system that students are in fact learning better? This seems to be the critical requirement for adoption of new teaching and assessment tools, i.e., for systemic reform.

The following remarks illustrate that Jorge Talamantes, Aaron Brower, and other writers also agreed with the members of the third panel that assessment and evaluation can be used as agents, or drivers, of change:

> Evaluation is the "missing link" between classroom practice and institutional practice. Institutions will not change without it. (Talamantes)

> Institutional assessment can drive institutional reform—if assessment focuses on "reformed" outcomes and can be done in credible ways. (Brower)

> Evaluation/assessment could be the means to make truly significant changes in educational policy; assessment justifies and gives credibility to our efforts to innovate. In addition, our assessment results should be used to influence others toward reform, including our national leaders.

> There is a difference between evaluation as an isolated exercise and evaluation as an agent of

change. Evaluations and assessment must be designed and written to communicate information effectively to the individuals responsible for effecting educational reform.

For a culture of change to take hold, it must be realized by all the relevant levels (professor, department, etc.) that change is necessary. Institutions are affecting that change through evaluations as feedback for appropriate levels of the educational process.

One participant summarized the insights of many with respect to the role assessment and evaluation can play as drivers of systemic reform:

> Where are we going? We must start by evaluating our current curriculum or programs, decide where we want to go, decide how we are going to get there (make a plan), implement the plan, evaluate the plan, revise the plan, etc., etc. Assessment should be ongoing, always subjected to scrutiny, and revised according to the new needs of students, faculty, society. If built in all levels—institutional, department, faculty, curriculum, student, it becomes part of, and a way of, life.

2.3 Student-Centered Learning: SMET for All

Some Forum participants focused attention on the need—particularly for those teaching introductory courses—to pay attention to the diverse backgrounds of students. They also wrote about the importance of responding to the needs of students with diverse learning goals, including: students not majoring in SMET fields, undergraduates preparing for further study in SMET fields; undergraduate and graduate students preparing for K–12 and college-level teaching careers; and students planning for employment in industry and business. These concerns appeared to culminate in the need to consider the purpose, or goals, of SMET curriculum for undergraduates:

> Students enter undergraduate institutions with diverse academic, cultural, and professional needs. How do we address these needs in preparing an undergraduate plan for their learning, especially for the major courses, which are content driven?

Barbara J. Tewksbury added her views on this topic:

> The market for Ph.D. academics will consume only a very small percent of the undergraduates that we teach. Paradoxically, that market is the only one for which many of us can articulate a clear list of needed skills and knowledge. Ninety-nine percent of our students will enter different careers with different skill and knowledge needs. What will be more important to them? I think we can make some headway just by being aware of that fact and tailoring our courses to make them most useful for the bulk of our students.

Many suggested that a shift from a faculty- and content-focused curriculum to a more student-centered curriculum was in order:

> We need to focus more carefully on student learning rather than on teaching and faculty. (Jim Swartz)

> If the dialog in higher education can shift away from just thoughts of what to teach to what students actually learn, this shift can further drive reform.

> Curriculum—are the goals of the curriculum student-oriented, or faculty-, or institution-oriented?

> If the paradigm is switched to a focus on "learning," assessment fits naturally: teaching involves a focus on context, a transfer of information; learning looks at what the learner constructs. Teaching puts the instructor at center; learning puts the student at center. Where we put our focus dictates what question we ask. So if we want ourselves and others to deal with issues of assessment, we need to put our focus on students. (Cathy Middlecamp)

> Any pedagogical changes should be made with students in mind, specifically their goals for the course. Change in assessment, change in presentation, should always coordinate with student outcomes according to the goals established. (Mike Gehner)

Many brought attention to the issues of diversity and equity in the classroom. Some wrote about the need to find ways to address individual learning styles and ability differences in the classroom. Some expressed concern that the "weed-out" feature of introductory or gateway courses may be unnecessarily exclusive and even counterproductive to promoting science literacy for all.

> I would hope that in the reform process factors relating to diversity and learning styles are fully taken into account and implemented in the changes. We are in dire need of true diversity in SMET, especially when you look at the trends in alternative education and means of understanding science and technology. Human nature tells us that we are different and therefore react differently. However, our educational system has little room for such differences, even in SMET. For example, Native Americans view science differently based on their culture!!! Student-centered teaching should take into account diversity issues! Hence the evaluations should be reflective as well!

> It seems that scientists (perhaps more than those in other disciplines?) have difficulty addressing questions of equal access to quality education. We often focus more on so-called objective forms of measuring students and deny that there are social/political/economic components to our choices about curriculum and instruction. It is difficult for me to see how we can have real institutional change without understanding equity or [unless we] see how our current assessments are used to sort students and determine access to future resources. [In other words, unless we understand the relationship between equity and assessment,] we will continue to choose assessment methods that help us do this sorting. This [lack of understanding] may cause us to cling to grading on content rather than process/understanding or to limit exposure to real science questions until later in the college experience (after the weeding process). (Laura Wenk)

> Alternative assessment strategies are necessary if we are going to move beyond the past participation types/results. They were designed to filter out everyone who doesn't look/think/feel like us. (Andrew Bernat)

A few writers also considered the impact of the curriculum innovations on students who thrive with the current curriculum and assessment methods.

> As acknowledged in the papers, our educational system produces top-quality scientists. . . . Do we risk sacrificing the development of these individuals while trying to motivate others? (Anthony J. Nicastro)

> One point not mentioned about changing assessment of students' knowledge and skills is that the door both closes and opens for different students. Students who excel on traditional assessment tools may do worse initially, but the flip side of this coin is that other students, who traditionally have not excelled, may now shine on these new assessments. (Tony Jacob)

2.4 The Value of Teaching

During the second panel discussion, the reward structure (particularly at research institutions) was spoken of as a major barrier to innovation. Think piece essays echoed this problem. Many writers described a faculty or department "culture" that places a much higher value on research and discipline area knowledge than on innovative teaching and assessment or educational research. These writers stated that the lack of emphasis on education permeates institutional and departmental support, funding, and reward structures.

Those few who wrote about their experiences with innovative teaching reported that their successful practices were

infrequently emulated by their colleagues and, often, ignored or devalued by both their colleagues and by the institution. Indeed, writers noted that many SMET faculty do not pursue innovative teaching and assessment practices because these are not valued and recognized. Even for tenured, senior faculty, undertaking innovation and education research can be risky because it may alienate the innovator from colleagues.

Also, many faculty noted they do not pursue innovation because of structural barriers of time, money, and resources. Some writers admitted that they lack the pedagogical knowledge necessary to improve their educational practices. They noted that more collaboration with education faculty and researchers could provide the resources needed to learn about successful teaching and assessment strategies. However, they also commented that many of their colleagues did not share the belief that SMET faculty could (or should) learn from education faculty.

Faculty Support and Reward Structures: Valuing Teaching and Innovation

As noted above, faculty must often contend with a departmental culture that does not value innovation and educational research. The following quotes by Robert Dehaan and Bruce Callen illustrate these feelings expressed by many:

> Only when provosts, deans, and chairpersons push for new pedagogy and new assessments, and the institutional reward systems needed to drive those changes, will individual faculty be relieved of the need to fight battles and accept risks that serve as barriers to change. The same argument can be made for accrediting agencies and national testing services. (Dehaan)

> I am very bothered by the idea that faculty should refrain from committing themselves to things they believe in and can make a strong case for. If the current system doesn't reward these activities, and we don't engage in them, I don't believe we'll ever develop enough authority to change it. (Callen)

Many wrote of the risk of pursuing innovation as being exacerbated by the tenure system. For example, one four-year college administrator wrote the following about risk:

> My experience reveals that research continues to drive tenure policies. This situation works against efforts to motivate faculty to seriously explore changes in teaching and assessment practices.

Junior faculty were seen to be at particular risk because the current tenure process tends to emphasize research and publication activities over teaching. Some people wrote about the need for senior faculty to support the efforts of untenured faculty:

> There are real risks to participating in innovations, assessment, or other [such activities] as a tenure-track faculty. Depending on institutional priorities for research vs. teaching, these may or may not jeopardize one's tenure. Too often, though, junior faculty face the rather unappealing choice of "maintaining the status quo" or innovation at the risk of poor student and peer evaluations (re: Mazur's analogy of the change in tennis grip). . . . I still bristle at the current state of affairs regarding the participation of untenured faculty in reform as too risky of a proposition. The early years of one's academic career are formative, or as Tapia referred to it, a "key transition point." There is consensus on this point. Most disturbing to me is an equally loud consensus that no (or at least very little) resources should be put forward with regard to smoothing the transition of young faculty into reform. (Stacey Lowery Bretz)

> Mazur said he does what he wants and doesn't bother to worry what his colleagues say. That might work for him, for he was a tenured faculty member at Harvard. It won't work for the untenured junior faculty member at an average college or university.

> Junior faculty are in the most difficult circumstances. At Portland State University, we have broadened our view of scholarly accomplishments in education. Nevertheless, junior faculty are generally encouraged by senior colleagues to play it safe and establish themselves as traditional scholars first, both to achieve credibility in their discipline and to maintain flexibility with their career options. (Carl Wamser)

> Untenured junior faculty should participate in these innovations. They come with fresh ideas. But the reward system does not allow them to take part in these activities.

> I would definitely like to see a change in the way faculty are evaluated. As a junior faculty member, promotion and tenure are very important. I agree with Richard Tapia that we have to be careful and protect ourselves until after tenure. However, those early years could be the most productive times to initiate change. It is an issue that I do not believe colleges and universities are aware of.

A few writers did not feel the reward structure inhibited junior faculty from learning and using any particular pedagogy:

> I see little sign that new faculty are in a risk position. They have to learn how to teach; any method they adopt will put them at risk, just as would any other including a classical position.

Some made the point that the current values are related to funding. Writers noted that teaching does not enjoy the same prestige as research because, at least in part, research attracts large funding dollars. As one wrote, "'Money talks' applies to universities, colleges, departments, and faculty." Similarly, Ann Redelfs wrote: "Pursuit of science is now tied to dollars, not to pursuit of science." One administrator indicated that faculty are pressured to align their activities with the values of their departments and institutions.

> Faculty will not so much take guidance, advice, and direction as they will respond to rewards or at least validation. If change provides them with financial, psychological, or other tangible or intangible rewards, they will pursue and implement that change. If change costs them more work or they perceive that there are no rewards, little will happen.

Consistent with these reports of the relatively low value most departments and institutions place on teaching were descriptions of overt resistance that a few faculty have encountered. One person mentioned the difficulty in "bringing current educational research into the classroom when there is opposition from other faculty and students." The following person who was involved in a classroom innovation described his experience of being rejected by colleagues:

> Newer faculty members now avoid me when I attempt to enter a discourse on learning. They say that their research is "too important" and they "don't have time."

Furthermore, it appeared that education research often is not valued by SMET faculty. The following people pointed out that some faculty tend to have less respect for education research and educators:

> Scientists and mathematicians are generally not inspired by educational specialists.

> All the speakers talked about need for serious, professional evaluation/assessment, but none addressed the cultural barriers involved between SMET and education faculty. Many of the former dismiss what their education colleagues have to offer as "too soft," or [said] there is a stigma associated with admitting one "needs help."

Some described structural barriers that faculty face when trying to incorporate new approaches into their courses. Many people noted that research and publishing expectations, along with the lack of resources and know-how that would help them incorporate new teaching methods, ultimately make the prospect of using innovations in the classroom daunting and

time-consuming. Faculty wrote about their experiences and observations:

> It is my observation that most SMET faculty are aware of but don't value [new approaches to] pedagogy. Though there are a variety of reasons for this, high among them are "time" or lack thereof. It takes time to inform yourself of the research base in effective SMET teaching and to plan classes accordingly.

> I am struggling with changing the ways in which I assess student learning so that they match the curricular innovations I continue to make in my courses. I have encountered the following problems in my efforts to do this: (1) time constraints make it difficult to give sufficient attention to constructing new assessment tools as well as to analyzing their results; and (2) my knowledge of how to construct these tools is limited. (Sharron Smith)

> In my own program I find that faculty who are of good will and who care about student learning [are nonetheless] reluctant to take even small steps toward small group/active learning formats. They want to keep lecturing because that is what they are most comfortable with. Other ways take too much time.

> To suddenly change to group instruction that promotes alternative forms of assessment places a great demand on time, resources, and faculty mindset. Many superior faculty see these demands as unacceptable when faced with issues such as research publication. Nontenured faculty see it as a great obstacle to survival at the institution.

> The first concern is the time cost of assessment in the innovative setting. My colleagues generally enjoy teaching out of the text such as the Harvard Calculus, but they are very concerned about the extra time needed to design, grade, discuss group projects. . . . Curriculum changes require time. My colleagues are evaluated on research and their score on traditional student evaluation forms. Most of those faculty not successful [on these criteria] were weeded out. Those left are comfortable. The push for change comes from the young faculty and college administrators who want a product they can take to the tax payers. The time commitment needed for my colleagues to change what they do efficiently and well is tremendous. (Gordon Woodward)

Faculty Performance: Making It Safe to Innovate

Numerous writers criticized the practice of using course assessment data for accountability purposes, citing this practice as a barrier to change. A few faculty expressed the fear that assessment data intended for formative purposes could be used to evaluate teaching performance and result in denial of tenure or loss of status within the department.

Bruce Callen and Deborah Warnaar were concerned about the use of course assessment and faculty evaluation data:

> My concern is that the [assessment] information provided by this process is also sometimes used for evaluation of faculty performance, which I believe is a terrible mistake. It destroys faculty interest and willingness to participate, since they are fearful that honest attempts to improve their courses will end up as arguments that they're not worthy of tenure or promotion. The result is that assessment becomes something driven from the top down, which creates even more resistance and hostility. (Callen)

> I worry about evaluation used for "punitive" administrative purposes rather than as an aid to further self-reflection and growth. (Warnaar)

A few others argued that faculty incorporating innovations are held to higher standards than faculty who use traditional methods.

> There is no evidence that the traditional courses are as "successful" as some cynics would claim, and they hold reformed courses to a higher and different standard of proof. (G. Earl Peace)

A few writers noted that faculty feel pressure to be positively evaluated within their institution. For example, Darlene T. Eyres and another participant wrote about one aspect of this issue:

> Many students who have traditionally done well may not get their customary grade when the emphasis changes in the classroom. They rate faculty low in the evaluation and the faculty revert back to what is safe. (Eyres)

> The summative student course evaluations inhibit change because the university attaches so much weight to them. In my program, those who give brilliant lectures get higher ratings than those who force students to struggle with the material.

Several felt that traditional faculty evaluations inhibit efforts to implement innovative practices because they pass judgements while innovations are still in the pilot stage.

> Comments about mistakes: in scientific experimentation, mistakes are tolerated. There is less tolerance for mistakes in teaching experimentation (student evaluations, views of traditional colleagues). How do we deal with this? (Nedah Rose)

> Departments and institutions can encourage change by such techniques as "hold harmless" understandings, where change in pedagogy, curriculum, and evaluations are allowed to interact recursively for a period of time without "summative evaluation." (Terry Millar)

Institutional and Departmental Support: Valuing Change

Writers made it clear that faculty concerns about the risk of pursing innovation cannot be ignored and that efforts must be made to encourage a supportive environment within the institution. As one participant put it, "Institutional support for change is critical [for change efforts] to have impact." Many called for a dialogue involving all organizational components. For example, one faculty member noted the need to

> generate more widespread discussion between faculty and university administrators about why reform is needed. A consensus needs to be reached about what type of reform is required, that systematic change is required, and how to share the cost (dollars, time).

Many Forum participants argued that, if the ultimate goal is to change the practice of teaching, then teaching must be valued at all levels of the institution, and this value must be reflected in the values, goals and practices of each level.

> Assessment is based on outcomes, so setting institutional, departmental outcomes—as outcomes supported at broader levels—is essential (in my view) as a first step. When outcomes support the mission of the institution, real change has a chance. Goal setting should be coordinated across the major, across programs, and across the undergraduate program (if feasible).

> From my perspective having played several roles (department head, dean, faculty member), there needs to be put in place at each institution a model of accountability in terms of outcomes. This is down at the course level—what we say is being done [needs to be] monitored and evaluated. Then this translates to department, college, and institution evaluations. Are all walking the same walk? Issues of how people are rewarded should not be dissonant with what they are to do in their various professional roles.

> Until administration begins to place more emphasis on the importance of teaching, true reform in SMET courses will not occur. The enforced emphasis does not have to be in the form of rewards, but must, at the very least, judge teaching as being on an equal level with research in the granting of tenure, merit raises, and so forth. (Nikki Privacky)

> Changing assessment techniques may help answer the questions on student learning, class content, and pedagogy raised by this

panel, but it does not deal with costs and risks unless the new assessments are accepted by the department (division or university) as valid. Doing this work takes lots of time already. Do we double it by also adding a major lobbying and public relations effort [in order to get] one's work accepted and valued by colleagues and administrators? (Sandra Lausert)

Many writers emphasized that different organizational components must provide reward and other support structures that are aligned with the new values. We need to cultivate incentives and recognition for those willing to change, wrote Nikki Privacky, Norma Davila and others.

> In particular, I would like to hear more about a topic that emerged only at the end of the forum; the payoff to faculty to make curriculum, pedagogical, and assessment changes. The incentives that were mentioned—increased calls for public accountability, pedagogical interest—are not always adequate.
>
> Without some form of recognition, there is no impetus for faculty to develop reform projects except that it provides them with intrinsic enjoyment. That is obviously enough for those of us at this conference, but for most faculty some form of external gratification is necessary. (Privacky)
>
> Expectations must be clear, but they are very much dependent on the institutional culture of where we are doing our work. Tenure can be a problem when the rest of the faculty does not want to support our efforts and could even misinterpret our own learning process. We need support for ourselves and our work to keep going. (Davila)

The following quote highlights certain types of support needed by faculty interested in incorporating innovations into their courses:

> It is important to recognize the particular factors that affect change at some universities/colleges. What affects one will not necessarily be important at another. Most critical would seem to be (1) the atmosphere and opportunity for change: Will a department/faculty/administration be receptive to change? (2) the commitment to change: Will the university/faculty commit resources to the effort (assuming that the need for change has been defined)? (3) Will the participants be "held harmless?" That is, will careers/promotion/tenure/ other negotiations be sustained and supported for those who are innovative or taking the risk? (C. Singler)

Some people also argued that community colleges and nonresearch universities enjoy an environment where education research and innovative pedagogy are valued and supported. For example, one person suggested that the risk nontenured faculty take when focusing on innovative pedagogy is more characteristic of research-oriented institutions:

> [The situation] of junior faculty: need protection or assistance from upper management at 4-year research universities. This is not necessarily the case at other types of institutions. For example, I was hired specifically with the provision that I participate in our teaching school system's reform effort.

Some offered ideas about how the system might respond in ways that incorporate both research and teaching values:

> An example [of a successful change model] would be looking at the dominant value of a department, which might be to foster research, and then at the specific goals and values of a small group of faculty in the department, which might be to foster students' thought processes and ability to investigate. The two could be addressed through incorporating real research into the course. Other values might be harder to reconcile, but we should search for these middle grounds.

3. Recommendations: Building and Sustaining Change

The first main portion of this synthesis focused on key needs, problem areas, and concerns about assessment and educational reform that appear in the participant think piece essays. Readers may have noted that the issues addressed in this first section restate and elaborate on many of the key points made by the panelists.

This second main section synthesizes the specific mechanisms, processes, and strategies for catalyzing change that think piece writers presented.[2] In this section, readers will note several important points made by think piece writers that were not made by panelists. Of special note, in this regard, are the points made in the "Communication, Collaboration, and Dissemination" section (pp. 53-55 below).

3.1 Catalysts for Change

Many writers described a complex higher education system, identifying both bottom-up and top-down pressure points that could be used to initiate and implement change. Writers advocated for processes to institute SMET education reform and make departments and the campus at large more responsive to student learning needs. They called for more professional development opportunities to learn from educators such as those in the K–12 system. They proposed systemic changes in communication, collaboration, and dissemination to make change possible and sustainable. They requested assessment tools, models of successful innovations and programs, as well as data on outcomes and "proof" of successful strategies.

[2] Because writers frequently presented needs and issues concurrently with recommendations, both "needs" and "recommendations" are presented in this section, resulting in some overlap with the first section.

Grassroots Reform

Some writers discussed the change process as one that starts with a few lone innovators, referring to Eileen Lewis's discussion of change that begins at the "fringe" and evolves as more faculty become involved. Some of these endeavors were referred to as "grassroots" change, implying that starting small and "growing" reform efforts from below was a practical way to effect change. Many writers conveyed their belief that lone innovators, or small pockets of committed faculty, could induce reform through testing and modeling innovative practices and spreading knowledge of successful practices to an audience of practitioners.

> This is clearly the key—with institutional support and national pressure, the "lone wolf" faculty in the departments will gain credibility and support.

> I tend to agree with Lewis, who says change can start at the grassroots level and filter up, to an extent. (Linda Tichenor)

> Personal experience as a martyr to this cause for the past 11 years says Eric Mazur has the correct implementation methodology in place. That is—do it without forcing it on your colleagues (especially within your department). Couple this with David Porter's philosophy—good teaching is like a virus, it spreads—and the result is a reasonable attempt at [managing] the logistics of implementation. One final important thought: nothing happens with no communication medium in place. (Jeff Turley)

Some however, voiced the opinion that grassroots activity does not necessarily lead to institutional change unless a high-level administrator takes the initiative. As Joel J. Mintzes explained:

> On my own campus, change has rarely come from the "grassroots" within departments, but instead has been catalyzed and sustained through intervention of "enlightened" (and

"benevolent") administrators. Most recently, this was seen when the chancellor provided funding for proposals on the use of technology in science teaching. Ultimately, interest in this effort spread as faculty began to view it as a legitimate enterprise.

Another writer questioned David Porter's "virus" analogy, suggesting that there is little hope that lone faculty innovators, or grassroots reform efforts, can affect the system:

> The assertion by David Porter that "innovation spreads like a virus" is much too global a statement. Viruses lower resistance (internally); innovation enhances resistance initially (externally). If, in fact, this "virus" were spreading, I should not see so many familiar faces at this Forum.

Before new practices are accepted and supported, cultural shifts within higher education institutions may be needed. Some people suggested that change in faculty culture would facilitate a change in the reward structure and other institutional norms that were portrayed as unfriendly to teaching innovation. Roger Nanes and Judy Ackerman gave examples of how participation of even a few faculty can provide successful pathways for spreading innovation:

> I believe that the greatest impediment comes from the difficulty of getting buy-in from faculty colleagues. My reason for optimism comes from the fact that, on my campus, I see more senior faculty acknowledging the importance of paying attention to pedagogy issues, and there is more involvement by these same research–oriented faculty than has ever been the case before. Ultimately, it is the faculty that write retention, tenure, and promotion criteria who will recognize contributions by junior faculty. (Nanes)

> In the department that has done the most, it has taken the persistence of a couple of faculty members who brought their ideas, offered professional development opportunities, stimulated brown bag discussion, and shared their classroom activities and assessments so it all became real. (Ackerman)

Articulation with K–12

Some suggested a change strategy based on learning from those with previous experience. Many at the Forum cited the knowledge that K–12 educators and education experts have accrued in the areas of pedagogy, knowledge acquisition and application, and assessment. They acknowledged that the K–12 education community has learned much about "what works" and "what doesn't work" through their experiences. Forum participants considered much of this knowledge to be applicable and transferable to the higher education community:

> I agree that K–12 has much experience in assessment that can inform higher education.

> [Not] until the university science, math and engineering professors fully understand the major changes that are occurring in curriculum, instruction and assessment at the precollege level, will they be motivated/charged to reflect on ways to change their own instruction/assessment strategies.

> Another thought reminds me that the pendulum swing in K–12 education is beginning in undergraduate education. That is, there is a swing away from the more traditional focus on content goals toward less traditional student goals (such as attitudes and skills) and toward pedagogy and method. The issue here is to learn from K–12 , where people are just now rediscovering the importance of content understandings.

Some focused on how collaboration with K–12 educators would help SMET postsecondary educators improve their teaching and thus student learning. A few highlighted how improving preservice teacher education in the science and math fields would improve K–12 teaching and facilitate articulation between secondary and

postsecondary classrooms. An administrator provided an example:

> I would suggest that university professors team with innovative elementary/secondary teachers to examine models of assessment that promote conceptual learning and critical thinking. Bridging this gap would not only provide mechanisms to enhance all aspects of science and mathematics preservice education programs, but would also provide excellent models for all undergraduate instruction. What is good for the goose is good for the gander.

The notion of a fully articulated K–16 education system—a seamless web of science and mathematics education—was promoted by several writers. They suggested that articulation was a two-way responsibility requiring communication and collaboration across educational institutions.

> There is a tremendous need for better articulation between the K–12 and university sectors and across the university departments that teach SMET content. The differing expectations and assumptions regarding content and process of teaching and learning among teachers and students in these sectors make assessment or innovation extremely difficult to implement or sustain.

Furthermore, it was suggested that a collaboration between science and math content experts and education experts would promote better teaching and learning:

> A second point is that we (as scientists) really need to team up with the "experts" in cognitive science and learning theory. We should not reinvent the wheel.

> Undergraduate science instructors could benefit from training in educational psychometrics and classroom social structures.

> A real issue is the disconnect between the expert groups: discipline-defined content experts, skills experts, assessment experts. There is a critical need to find a way to build bridges between these groups.

> How can we convince math/science professionals to work with education professionals in setting appropriate goals for the math/science education of future teachers? These types of discussion, and collaboration, must precede efforts to modify the evaluation and assessment of courses for these majors.

Professional Development: Learning about Teaching and Learning

There was a call for more resources that faculty can use to learn about curriculum reform, assessment strategies, and innovations in teaching and learning. Often this need was linked to the recognition that many faculty have little formal training in, and have not developed proficiency in the practice of, teaching, pedagogy, and learning. This recognition led to a variety of requests for professional development for faculty, graduate students, adjunct instructors, and administrators:

> The majority of the Forum attendees are neophytes. This indicates we need more opportunities to work through how to identify appropriate curriculum goals, and how to implement tasks and assessment tools.

> Professional development is still a need at all levels. Knowledge and expertise are needed as well as the commitment and will to change. How do we help develop tools for use? What professional development is needed to help the masses use them?

> To facilitate SMET faculty to use a variety of teaching and assessment strategies in their teaching, they need to not only know what they are but to value pedagogical content knowledge as a viable tool to facilitate learning.

> Too little attention is given to good assessment coupled with good instructional design, both at the individual course level and at the program level. Little is available in the way of professional development activities with follow-up mentoring. Even if such resources were available, heavy teaching loads at community colleges leave faculty with little

time to develop good assessment and evaluation processes, practices, and policies.

There was little discussion of the role of graduate students as a vehicle of institutional change. They are the institution of the future; training them in new perceptions of teaching sows the seeds on 10-year time scales for deeply rooted change. These seeds may lay fallow through postdoc/assistant faculty years, but they will not die. (Robert Mathieu)

Examples, Indicators, and Proof

Forum writers believe it is essential to have examples of successful innovations, models and tools for implementation, as well as established indicators of success and data that provide evidence of that success. Time and again, writers requested examples of assessment tools, "how-to" specifics, and documented models of classroom innovation and institutionwide reform. Faculty in particular called for specific examples of new practices, examples of actual assessment tools used in the undergraduate classroom, and more resources for developing and implementing different assessment measures. They want valid and reliable assessment instruments for the classroom and quality evaluation of programs.

For years we've heard that in order to assess student knowledge and performance, we need to identify goals and objectives. We need some tangible examples of how this is done in life sciences.

Modeling is essential if there is to be a substantive change in the knowledge of assessment in future professors and teachers. (Lars Helgeson)

The popularity of the University of Puerto Rico assessment plan (in terms of the number of individuals seeking a copy) implies to me that there is a great need to take successful models of institutionwide reform and present them as national models.

I believe that building a library of successful examples of such congruence will be essential in expanding the community of people involved in this task. (Bruce Callen)

Perhaps it is important here to focus on a collection of "examples," practices of transitions: How was this accomplished? How were common goals agreed upon? How was acceptable evaluation/assessment developed?

We should identify models [that are] effective in bringing about desired change. Try out these effective models in selected institutions, communicate both "raw data" and "meaning" of the findings.

A number of Forum participants wrote of the general need for more information, data, and research on innovations and assessment practices:

We need to do a better job of documenting how we know what we can claim to know. We are in the habit of asserting what works without supplying the research evidence to support it. We must begin to ask the same questions about how students learn and how we know the impact of instruction that we would ask about our scientific research.

The most useful thing we could provide for institutional change (including school and departmental change) would be clear, concise and valid data that new approaches to education are as effective or more effective than traditional approaches. (James Highsmith)

Scientists are empiricists. It is important to provide them with scientific data that support the kinds of reforms that we have been discussing because "seeing is believing." . . . the laggards and the cynics will not accept the need for change until they are shown that their preconceptions are wrong and that other models do work. If this sounds like inquiry-based science education, it is. What we prescribe for our students is probably what we should be subscribing for our colleagues—learning (about new approaches) by doing, and having their preconceptions (about lecturing

and about diversity of learning styles) challenged. (Ken Verosub)

Both administrators and faculty wrote about the need for documentation and data that provide evidence that innovative practices work. They want clear, concise, and valid data about proven practices in order to persuade faculty to try new approaches:

> In addition to coordinating outcomes with the mission of the institution and examining resources required at the outset, assessment design, evidence collection, and interpretation seem critical in sustaining change. (This should require more than anecdotal evidence and speak to original goals.) Without convincing evidence (i.e., task change has increased learning, and met other objectives), that particular innovation may be regarded as ineffective.

> Coming from a department in the College of Sciences, quantitative assessment data would be some of the most convincing evidence to support change in course delivery—once total information is generally compelling. If we could support new teaching/learning methodologies by statistical comparison, it would be much easier to recruit more faculty for innovative course offerings.

Moreover, some wanted common measures or indicators, so that data can be compared. To this end, they wrote that development of baseline data, common indicators, measures, or national standards would provide for consistency, comparison, and quality information.

> From the national perspective, it would be extremely helpful to have a set of common baseline data so that innovations could be compared across a variety of institutions and departments.

> We need to develop and implement measures of student and system achievement. It would help if a consensus were developed on all levels of the system—similar things have been done for the K-6 educational system. (Kevin Aylesworth)

> In order to develop assessment tools that are informative and useful across curriculums, it is imperative that we develop common national goals that would exert a powerful influence over our academic institutions.

> The panel did not address the difficult issue of how one can aggregate across all the individualistic inventions in assessment of SMET courses developed by individual faculty to develop national indicators of student learning. That was the challenge thrown out by Luther Williams—and it is one that is unlikely to go away. (Senta Raizen)

> When we adjust the assessment to the creative teaching, what do we use for comparison? We must be able to prove with comparison and with another group that we are doing well. (Zafra Lerman)

Some affirmed Gómez's point that there is a need to develop measures that address the concerns of diverse stakeholder groups—faculty, administrators, as well as national agencies and public audiences:

> Evidently [we] need diverse types of data that can be accepted as "information" or evidence by the different audiences. Faculty are more likely to be interested in student learning outcomes (or should be), while administrators are going to want information on efficiency.

> Translate assessment process and products into meaningful/useful/compelling terms for the varied stakeholder audiences: Find "key" footprints that generate "Aha's" to maximize engagement and use. [As Gómez put it] we need to be great in interpreting and outstanding in communicating.

> The "public" seems to be demanding more external evaluation, not less. If our reforms are to succeed, they must produce results that are valued by society at large. (Marlene Moore)

"Assessment data should be treated with the same rigor as scientific data" wrote one faculty member. Such calls for validity and reliability of assessment data, student record

data, and evaluation data were common. Similar views were expressed regarding the need to ensure quality assessment tools:

> If "alternative," "nontraditional" assessments are to convince and transfer to our colleagues, they must stand up to reliability and validity criteria. We must assess the assessments!
>
> Assessment in this case must include assessment of the evaluation criteria and the instruments themselves as a means to enable the institution to be self-critical. (Herbert M. Dyasi)
>
> Assessment is very important for improving teaching and learning at all levels. The tools used for assessment must also be constantly upgraded as new problems and knowledge evolves.
>
> Assessment should be ongoing, always subjected to scrutiny, and revised according to new needs by students, faculty, society.

Furthermore, writers conveyed the need for evaluation feedback to improve courses, and foster achievement of stated institutional goals:

> Faculty need evaluation results in order to make more informed decisions about how they might improve their teaching.
> Perhaps the message needs to be "innovations followed by institutional support for innovations that have demonstrated a benefit to the students." If the innovation was not clearly a benefit, the follow-up questions (Can we adjust it to make it work? or Should we forget this as inappropriate for our situation?) need to be asked and clearly answered. We simply must take the time to evaluate. (Charlotte Otte)
>
> In a way, assessment and assessment results can be objects of study to judge what are the institution's goals and values, and if these in fact match the stated and implied goals and values.

Communication, Collaboration, and Dissemination

Communication, collaboration, and dissemination were all addressed as key strategies for furthering innovative practices and improving assessment strategies for many who wrote about promoting systemic education reform. Communication was seen as vital both to begin and to maintain the conversation about systemic reform. It was also seen as a way to pass information, change beliefs, and motivate faculty and administrators alike to step outside of the traditional mindset. Collaboration of a "critical mass" of participants was portrayed by some as another foundation piece upon which to build reform. Many suggestions were made on how to improve dissemination of information, data, and specific models and examples. These included requests for publications, web-based materials, implementation resources and assistance, libraries, professional development opportunities, and national workshops and conferences.

Many participants wrote that communication, between and among both those practicing innovation and those interested in learning more about innovation, is essential for engaging in and sustaining change over time. Both Laura Markham and David Bauman made this point.

> One of the keys to implementing change is the dissemination of ideas. Sometimes you find that you are reinventing the wheel. We need to see ideas/research flow more readily from education/cognitive science to and between the disciplines. There is a small community of, say, chemistry education researchers, but practitioners need this sense of community also. We need substantive conversation to flourish. (Markham)
>
> Communication between and among all participants (K–16+, business and industry) is critical. This relates to expectations for students and by students. Coherent, well-communicated linkages are critical for

changes from what business needs, to course content, to how classes are taught and assessed—and this includes how students learn. (Bauman)

Scale-up will require collaboration and consensus among a critical mass of committed individuals throughout the SMET community.

Change is a complex process that requires coordination and broad consensus. (Eleanor Siebert)

At present, we at Madison Area Technical College have pockets of activity and commitment to change, but institutional reform will wait until the committed group becomes the critical mass. (Joy McMillan)

I would echo the speaker this morning who focused on the need for increased and more effective communication of assessment data and outcomes, especially from the point of view of the faculty to administrators. The major challenge of institutional change is convincing those with power to use that power to the advantage of the reform/assessment process.

Communication efforts need to engage and inform a variety of stakeholders in the public at large as well as inside the SMET community.

Can we engage the broader community in conversations about educational goals for all students, for all citizens? Can we engage prospective "employers" (at whatever level of specialization they may require) in helping us communicate to students the types of performances that will be valued in their professional lives? How can we encourage and enable K–16+ articulation so that teachers support each other in risk-taking and innovation toward more effective teaching and learning? (Judith Pelchat)

A change in the reporting of assessment is needed. This reflects a need for enlightening the public about the purpose of assessment so that schools will have new expectations to meet. (Lars Helgeson)

We must include community and legislators in the conversation to affect change, as politics often hinders progress.

In light of the funding discussion: change requires the cooperation of the larger community. We must educate the larger community to the value of change. Also, it is the responsibility of the larger community to support change.

Several writers suggested that change can be fostered by promoting collaboration, including common professional development activities, among diverse stakeholders.

Collaboration among disciplines (SMET, education, psychology, sociology, public policy) is possible at all campuses. These collaborations encourage institutional dialogue, conversation, and change.

With my belief that K–16 communication is a key component for sustainable change, there needs to be a facilitation of communication. I feel this can be accomplished through common professional development.

A large concern is need for even a small community of like-minded, invested colleagues to generate our own approach to our "professional development"—reflection on and improvement of the approaches that we design. This is very difficult when the focus of assessment is fragmented by different levels of the educational process. (Nancy Carnal)

Forum participants wanted information, resources, examples, access to expertise and opportunities for interaction with experts. Writers suggested a number of dissemination strategies to get the word out, and to get help to those who need it most. They suggested publications, libraries of resources, web-based strategies, and national workshops and conferences where successful practices and tools could be shared and distributed.

There need to be both measures of student outcomes and places to publish and disseminate information about reform efforts (i.e., efforts to improve student learning) in order for most faculty to agree to a balanced model of evaluation.

We need to develop more opportunities at the national level for presentation and discussion of indicators of success. (Kylie Keshar)

Some even had suggestions for specific dissemination formats.

The workshop model for dissemination of innovative ideas has worked very well for the geosciences. (Barbara Tewkesbury)

A role NSF could play would be to organize regional (state) conferences that bring together the universities /colleges wishing to systematically change the evaluation process. . . . Along with these institutions, accreditation organizations would also be invited. The conference should not be a series of presentations, but rather should have a structured, task-oriented focus. (Tony Jacob)

There must be a forum for the collection and dissemination of the results of assessment.

One suggestion is that NISE include this on a home page-chat room—and take best suggestions.

NISE could do a lot to promote change in the existing paradigm by supporting young faculty's efforts through recognition. Indeed, we would all do a lot for the cause if we each went back to our respective institutions and just planted a seed of change or brought these ideas/attitudes to faculty meetings where new hires are made and tenure decisions rendered. (Peter Dorhout)

3.2 Sustaining Institutional Change and Fostering Systemic Reform

For many attending the Forum, institutionalizing changes in assessment, evaluation, and innovative teaching practices means comprehensive, sustained reform for undergraduate SMET education. It is not enough for improvements to be made in a few classrooms and colleges. They must be made nationwide.

We need to promote dialog and shared vision among faculty, administrators at the institutions of higher education, and state/national level policy makers around the global issues of attainment of science literacy and overall institutional efficiency.

Many wrote that while higher education institutions must often rely on external funding and national leadership support to initiate improvements in education, the responsibility to sustain and institutionalize change efforts ultimately also rests at the institution level. Nikki Privacky and Warren Hein wrote:

It is not up to the NSF to fund all SMET reform. The education institutions must play their role as well. For even with NSF funding to begin reform, sustained efforts can only be supported by the institution. (Privacky)

I agree that, if there is going to be a change in the way we assess our educational experience, there must be some support and initiative at the institutional level. (Hein)

Values and Policies

Participants believe that institutions must use both policy and practice to create an environment conducive to change. They must make changes in the way values are articulated daily on campus and in the way these values are supported in reward and other polices. Many recommended that a common set of institutional values be developed to successfully promote campuswide goals and support a culture of change.

We need to create a climate for change and a process to create, challenge, explore, reflect, fail, discuss, and succeed.

I was most intrigued by two concepts brought up in these discussions. The first was cultures (the environment wherein change must occur), and the other was receptivity to change. The value systems imposed are critical to the development of a culture where change is possible and nurtured. (Robert Weinbeck)

Although many assessment undertakings may be driven by the external environment (e.g., ABET), the key to successful assessment and evaluation is the value placed on them by those at the top. If they are valued at the top, resources will be found and faculty will be committed to these goals. (Alberta Lipson)

Others recommended that collegiate institutions as a whole must not only value change goals, but also must establish and implement policies that reward the change effort.

Distinguishing between situations that discourage, permit, or foster change may be useful; moving from one to another of these levels of involvement will usually require conscious decisions by the unit involved (and those above it) to value and reward the desired behavior. Instituting change may require multiple strategies for differing settings, and institutionalizing change may require as much conscious effort as initiating it. (Brock Spencer)

This is a "system" problem, and we must begin to identify the pieces of our system, define them, assign value and reward to them, and promote understanding of the system process and approach. The infrastructure for assessment and evaluation should be "buildable," once we understand the scope and breadth of all the system pieces. Sharing knowledge about indicators and measures is very important. Finding common measures which the entire system can use—Dr. Gómez's macro variables—will help everyone who is approaching the assessment and evaluation arena with confusion and trepidation. (Ann Igoe)

Some emphasized the importance of establishing institutional capacity to use assessment and evaluation findings to improve programs and disseminate successful changes.

Interpretation of assessment data at the departmental level is critical—what is done with the data can influence change in the community (both positive and negative).

Often only those directly involved in large-scale evaluations (e.g., accreditation) are ever informed of the results, and there's rarely any thoughtful discussion (at departmental, college, or institutional level) about how to use the results to improve the program.

After a few faculty initiate change, it is crucial to provide the resources to export this to other faculty who express interest in buying in. (Adrienne Kozlowski)

Some writers promoted the idea that institutional structures must be designed to accommodate change, and that institutions must have a process for change that identifies a starting point, goals, resources, and outcomes.

The broader organizational structures in which undergraduate education takes place are part of the puzzle of improved assessment learning. Structures need to be supportive. In addition to coordination in setting goals, consensus needs to be reached regarding resources required. If institution, department, etc., are not willing or able to sustain the changes needed to achieve goals, some adjustment (or marketing) needs to take place

Manipulation of resources means an examination of institutional values and aligning practice with those values. The value are contained, in part, in assessment practice and also in other dimensions (e.g., promotion, tenure). (Herbert Dyasi)

Evaluation is only effective in promoting growth and change if it has institutional support. There must be institutional goals which are accepted down through the department level.

Funding

To a certain extent, all higher education organizations, whether two-year, four-year, or graduate level research institutions, are affected by, or reliant on, external funding sources. Writers discussed the various external funding agencies that can influence change and be leveraged to initiate and sustain change. Others voiced a note of caution, reminding their colleagues that money often comes with strings attached.

A few writers were acutely aware of the role external funding can have in determining the nature of research, teaching, and learning on college campuses. Jacqueline Haynes wrote of the external funding agencies' influence, noting that their power is often underestimated:

> The role of funding as an extremely powerful force in establishing the research agenda, assigning priorities, and providing a reward structure seems to be underestimated by the funding agencies. The notion of significant research programs taking place in the absence of this funding is a myth (today at least).

Another participant identified a downside to reliance on external funding sources:

> External incentives can leave an effect which is both insidious and pernicious. They can all too easily become the tail that wags the dog. They also drive the system forward, creating appearances and flashy images rather than systemic changes of significance.

A related concern about the dependence of higher education institutions on external funding was characterized by the question, What happens when the funding runs out? Two participants wrote of the need to anticipate the next steps to ensure longevity for successful initiatives:

> In light of the funding discussion: . . . It is a reality that, once funding has run out, institutions fail to continue the support to keep the innovative programs running. What can we do? Maybe we need to spend more resources on effectively educating our ongoing-fund sources (state, college) to the value of change, new programs that have shown success, assessment in new ways, etc.

> Coppola's comments about programs that receive funding, yet fold or die after the funding cycle ends, also tie in with the value system held in many of our institutions. If the programs never become self-sufficient, then they will forever be dependent on outside funding. This raises the question as to whether the program actually exists to implement change or to receive and secure funding.

Despite these concerns, external money was also portrayed as a positive resource to initiate change. As Cathy Middlecamp put it, "Often, quick changes are catalyzed by some event outside the system (it is hard to push a bus when you are on the bus)." Similarly, Raymond McGhee, Jr., suggested, "Outside money (NSF, etc.) has a role to play in catalyzing change."

Some had suggestions about how to improve allocation of external funding sources for maximum impact. It was suggested that "funding should be channeled to programs that are likely to succeed and have broad impact." Other writers outlined additional funding priorities for funding agencies.

> Adopters should be rewarded in addition to innovators. For example, funding agencies should make funds available for implementation in addition to development.

> NSF sees itself as a catalyst and can give its limited resources to those with the most potential for sustaining innovation. Given this, it seems important to move away from funding isolated projects and fund more projects at an organizational level—across departments and involving multiple levels in the organization.

Carl Wamser also credited professional societies with the power to influence teaching

and learning within disciplines and through funding and national prominence.

> Most faculty look to their disciplines for ideas on what is valued; faculty think of themselves first as a chemist, second as a professor of chemistry. So the American Chemical Society sends strong messages regarding what chemists should be doing based on the awards they give, the funding they support, and the media attention they focus on specific activities.

Accreditation

The role of external accreditation bodies and professional organizations captured the interest of many writers. Some writers considered these organizations as having the potential to catalyze change and lead SMET education reform at the national level. Some referred to accreditation as a strategy for establishing common outcomes or indicators to help instructors determine the "right footprints in the forest" and nationally relevant variables that can be used to drive the higher education system. For example, one participant wrote, "While the panel didn't really address the topic, I feel that national bodies (e.g., accrediting bodies, ABET) can drive policy and practice change nationally." Others described a more limited role for these organizations, depicting accrediting bodies and professional societies as potential partners in shaping institutional change.

Some participants, however, were not sure that these organizations would have a positive effect on change. They recalled negative experiences with past associations and referred to accrediting bodies and professional societies as mere "bean counters." For example, one writer commented that, often, faculty and departments are resistant to assessment and accreditation reviews even though assessment information can be used for improvement.

> Our university is currently mandating program assessments as part of the North Central Accreditation process and many faculty are quite resistant. They see it as a burden rather than as anything that might be useful. Yet, the first step in designing an assessment plan is coming to consensus on the goals of the program—that in itself is a powerful way to begin the process of improvement.

Whatever their regard for these organizations' past roles, many acknowledged that the potential to leverage these organizations as resources for affecting change was significant. Some claimed that professional societies could help determine standards, defining what is valued within the discipline in terms of what students need to know and be able to do. Similarly, accrediting bodies could provide a framework for measuring and evaluating institutional, departmental, and faculty practices by disseminating effective evaluation tools and examples of successful models. One Forum writer noted that such improvements were already emerging in the accreditation process:

> I am familiar (too familiar) with ABET accreditation, having been through accreditation four time using old bean counting criteria and once with trial new criteria. The current new criteria are much more open to local interpretation which is based on the local program and its own measure of success. However, it also gives a local campus a broad range of ways that they can adapt to the criteria—writing goals and developing means to determine how these goals are met. A great advancement.

Leadership

Writers noted that leadership is critical for the change process to succeed. Leadership is needed to initiate funding, ensure collaboration, and channel and allocate resources. Recommendations for leadership were plentiful, with writers suggesting

diverse sources for direction, guidance, and management. They wrote that professional societies, national agencies such as the NSF, and accrediting bodies need to take a leadership position in inspiring change and developing pathways that lead to sustained efforts on a national scale. They also noted that leadership support is needed at the institutional and departmental levels, among administration and faculty alike.

Some participants emphasized that leaders are needed at the local level and suggested the need to nurture leaders among faculty.

> The solution requires college administrators to build cohesive programs, departments, and divisional units with key faculty leaders. These faculty may emerge naturally, but in many instances, they must be recruited and restructured.

> Much of this can be accomplished by grassroots efforts that attempt to bring faculty together from various institutions. The efforts such as those in Florida (Higher Education Curriculum) begin to serve as a mechanism for discussion, investigation, and change and, perhaps most crucial, [to provide] support for those who desire to implement new ideas in their classrooms. (Nancy Romance)

Two writers described a role for NSF in guiding assessment and evaluation on college campuses.

> It seems like this is one way NSF (and others) could lead evaluation, by helping people learn enough to know who to talk to and where to look for the collaborations on their own campus and what principles they can use to guide the process of developing their own assessments.

> NSF could push accreditation of colleges to assist with institutional change.

Professional societies, accrediting bodies, and colleges of education as well as policymakers were touted as potential leaders. David A. Bauman and Robert L. DeHaan made the following suggestions.

> While this Forum is critical, policymakers must be participants to influence (encourage, motivate and reward) the necessary changes. Colleges of education must become leaders in teaching and assessment. (Bauman)

> Until (a) promotion and tenure decisions and other rewards, (b) pressures for fundable research projects and (c) national testing programs (MCAT, etc.) are altered, departmental faculties may be effectively frozen in the status quo. Thus, efforts by the NSF, national scientific societies, the national academy, and the education leadership might well focus on these higher level changes in addition to programs aimed at individual faculty and departments. (DeHaan)

Another writer recognized the NISE for assuming a national leadership role.

> This is clearly the key—with institutional support, national pressure, the "lone wolf" faculty in the departments will gain credibility and support. It is time to get beyond anecdotes and individual stories. Maybe NISE is in the early stages of doing this.

Implicit in the think pieces of many participants was the concept that leadership is essential at the national level to articulate the big picture and a coherent direction for change. Some writers were concerned about piecemeal efforts, suggesting the need for these efforts to coalesce into a larger, nationally focused initiative through leadership. Many at the Forum described the need for national leadership with a note of urgency:

> Clearly leadership is needed, and this Forum and the efforts of the NSF and NISE are progress, but it is time to get somewhere.

> Who will pull together the results of studies, experiments, pilots, etc. at institutes of higher education and then have the funding to

present this information to an audience of faculty throughout the U.S.? (James Highsmith)

Some expressed hope in emerging national collaborations:

> I see movement on the part of the content-focused national groups, such as American Association of Physics Teachers, toward assessment tools that give information on student learning. I'd like to see more of an alliance between the work of AAPT in assessment with NSF and NISE. This could strengthen the ability of individual instructors to bring data from such things as Hestenes Force Concept Inventory into assessment strategies departmentwide, instead of just within the classroom.

In sum, participants demonstrated through their writing that the Forum stimulated new ideas, reinforced key features of individuals' own experience, and motivated the framing of suggestions and plans for action. The test of the value of this Forum will be the changes—both subtle and major—it stimulates in policy and practice at all organizational levels. To improve postsecondary SMET, it is essential that the dialog fostered among diverse stakeholders at the Forum be moved forward, and that this dialogue motivates and helps to sustain infomed change.

Reflection and Concluding Remarks

Closing Speakers

Cora B. Marrett
Vice Chancellor for Academic Affairs and Provost
University of Massachusetts, Amherst

In offering a wrap up, I'll consider this Forum with the NISE vision as our framework. That vision—to set the stage for more productive science, mathematics, engineering, and technology education in the United States—puts the focus on productive education. Productive education for NISE, I would imagine, converges with what Diane Ebert-May describes as active learning, evident in understanding, reasoning, and utilization. Productive education quite likely generates the excitement in what Eric Mazur envisions as a higher quality of life for all, not only for those who become scientists or engineers. Productive education most certainly draws in the sensitivities, the understandings, that Richard Tapia recognizes as too often lost to science, mathematics, engineering, and technology because of untapped talent.

NISE then lays out the agenda, proposing we are engaged in the quest for productive education. That requires, of course, assessing where we are currently. The view that we must move to productive education implies that SMET education is currently less productive than it might be. But NISE is too firmly rooted in the empirical world to accept an assertion of nonproductivity or to endorse so sweeping a generality. My former colleagues do not rest simply on assertions, and thus they must have thought of convening such a Forum to ask, What, in fact, do we know? There must have been a set of driving questions that prompted them to organize the Forum as they did: What examples exist of productive education? How do we know that particular approaches produce given outcomes? What are the indicators of nonproductivity? Do those indicators derive from assessment of what students know and how they know it? Are the indicators the consequences of inquiries that demonstrate the value that specific content or pedagogy will add? Again, what do we know and how do we know it?

There must have been a recognition on the part of those who organized the Forum that a movement to productive education requires attention to assessment and evaluation. These two—productive education and assessment—deserve joint attention for at least three reasons. First, what students, faculty, and others produce, since we are talking about productive education, is likely to be shaped by what is measured. Consider the observation from David Porter that every system is perfectly designed to yield the results observed. If nonproductivity is the result, is it the measurement system that is the culprit? We have a second reason for looking at the interplay between productive education and assessment. We seek efficient as well as effective strategies for achieving productive education. I think my colleague, John Wiley, will join me, as we must as administrators in higher education, in reflecting on the problems we face with limited resources. Were the resources unlimited, we probably could invest in any strategy that achieves productive education. With the limits, the question becomes, What strategies prove effective and yet efficient? The question requires an examination of the several roots to which we can move in thinking about productive education. Assessment and evaluation then become invaluable for charting the course. There is

also the connection between productive education and assessment, for in the world we inhabit, the world of scientific skepticism and enhanced accountability, evidence counts. It is not enough to claim that productive education can occur; clear and convincing evidence matters. Thus I see the Forum as being organized to help assemble that clear and convincing evidence. There will have to be evaluation as we are thinking about assessing what people know and how they know it.

In trying to think of the connections between productive education and assessment, I'm amazed, not surprised, that my former colleagues again reached out to draw on extensive expertise. NISE chose to draw on a cross section to share knowledge, to verify assertions. We have the diversity of institutions represented: community colleges, comprehensive and research universities, four-year institutions, government agencies, professional associations, consulting firms. I conclude that NISE wanted no lessons overlooked and no generalizations overdrawn. Note, too, the specialties here assembled: students of the physical, biological, designed, and social world; experts on educational evaluation; publishers of books in science, mathematics, engineering, and technology. My assumption is that NISE understood that complementarities might exist, that specialties potentially can support rather than supplant one another. The goal is not to make everyone exactly the same, but to think of the intersections in the way in which assessment would drive and advance productive education. Finally, note just how diverse the responsibilities are of the people here assembled, from classroom teachers to program managers, to researchers, and, yes, even administrators. From their array of voices come some consistent themes.

Let me reiterate three themes that stand out from the general sessions, the plenaries, and the discussion groups. The first holds that goals for SMET education and assessment strategies must be aligned. The insistence and demand for alignment would occur in observations such as these from the various small groups:

- It is important to document goals for student learning and determine where we really want to go.
- It is necessary for students and faculty to have a clear statement of measurable course goals and to fit these in larger curricular goals.
- Good assessment begins with a clear vision of what the outcomes for learning should be.

These voices caution against disembodied measures or disembodied indicators. They suggest we should measure what we value and make clear the alignment between the ends sought and the measures applied.

A second consistent theme is that assessment should be approached systemically, in two respects:

- Assessment of the educational enterprise should be broad based, including analyses of faculty and services, as well as of student performance.
- Curriculum, instruction, and assessment must be integrated and seen as a single whole piece.

Earlier today at the plenary session we heard a call for aligning accrediting bodies, performance measurement systems, and systemic reform initiatives to effectively promote the productive education we have in mind. We heard as well some of the challenges that exist as new systems come into place; the virtual university and the high rates of movement across institutions muddle somewhat our traditional notions of system. What will we mean by systemic analyses with these matters that we must take into account?

But there is still another consistent theme from the discussions: Faculty require support, incentives, and justification to change assessment practices.

- Faculty need mechanisms, tools, processes, training, to learn how to apply assessment and learn what it can do for them.

- There is a need for specific models for courses of instruction including, for example, how to create active learning environments.
- Create a new system that rewards good teaching.

Time and again, participants have maintained the importance of the support as well as the justification if there is going to be alignment between assessment and productive education.

There are other consistencies. But I would be remiss were I to suggest that everything ended up being totally consistent. A few inconsistent observations emerged from discussions. I will mention only the question of transferability of experience across levels.

First I provide two observations suggesting that we have much to learn from what's taken place at the K–12 level.

- The K–12 community has been working on assessment for a good fifteen years. Higher education should learn from these efforts.
- The assessment vision is better articulated at the K–12 level than at the college level.

A third observation reminds us that not everything ends up as a matter of consensus.

- All issues, particularly assessment, seem to be more difficult in K–12 than in higher education.

Yet, there were relatively few such inconsistencies in the discussions, possibly because of the nature of the discussions. Could anyone argue against these statements? "Communication about assessment should be to the public and should be two-way." "It is important to look at the process of change and at issues of adaptation and dissemination." Surely they are not statements we are going to argue about. Thus, sometimes the absence of inconsistencies, or the fact that things seemed to converge, had to do with the very normative, abstract kinds of comments that people made.

Let us consider what must have been the NISE agenda in framing this Forum. If NISE expected systematic reviews of efforts to assess against particular goals, then I am not so sure that this end has been clearly achieved. We did not work through, in a systematic way, what the research demonstrates or what people actually understood about their own experiences. Against that kind of standard then, perhaps there is more to be done. But I would propose that NISE had other aims as well. NISE, I would surmise, wanted to know whether a community of committed specialists considered assessment a subject worthy of closer examination. In other words, is assessment a problem that is worth our time and attention? Has something significant been identified?

I would suppose that NISE also wanted to know whether the interest in assessment would be sufficient in breadth, depth, and potential to engage a range of communities, engage them in the thorny problems and rigorous analyses that would apply in context-related fields. Is this something engaging enough to bring to the table a number of different communities? And I suppose that NISE had in mind asking whether there is a will to probe the individual, institutional, and national conditions enhancing effective and efficient assessment and evaluation. These conditions would insure that assessment and evaluation do not become ends in themselves, but are signs of our accomplishments as we move toward the productive education and the enhanced participation in the society to which so many of the presenters addressed their comments. Against these kinds of standards, there is reason to be very optimistic. Responses signaled not just an interest in the subject of assessment, but a view that assessment becomes an important window on science, mathematics, engineering, and technology more generally. This topic, people seemed to say, is not only important, but it is a good way of moving into

the question of how we cultivate productive education.

Perhaps however, it is too early to evaluate the Forum. When we look at education, we often look at outcomes over time, not just what happens on a day-to-day or semester-by-semester basis. If the long term is the basis on which we do our examination, then there are questions that we will have to ask about the Forum later. There may be some outcomes not noticeable right now. I would imagine there will be an acceleration of activity by the College Level One team, in part because of the number of hits on the Web site that will occur now that people know what's possible, but also because of the desire to build a community and the attraction of the idea of an Institute[1] within an Institute designed to enhance our understanding, our improvement of assessment for productive education. But I would also imagine, and perhaps this is the dream, that from these activities will come the forging of communities—communities on campuses, in regions, around postsecondary SMET education and its assessment. I do not think that NISE or NSF had in mind simply having all of the responsibility rest with a single organization. From the Forum, great potential exists for the kinds of collaborations that will move the process forward.

In closing, let me return to the NISE objective. The Forum, as I see it, clearly addresses the objective of promoting productive education. I conclude that the aim of linking assessment to productive education is, in the imagery of David Porter, a reasonable vision and not a hallucination.

John D. Wiley
Provost and Vice Chancellor for Academic Affairs
University of Wisconsin-Madison

Last evening Cora and I met with others who were helping us pull out some consistent themes that we had heard at this Forum. Our overall conclusion was that, to the extent that this Forum was designed to provide a launching pad for the College Level One team, which is going to spend the next year doing an intensive examination of assessment, it succeeded brilliantly. We really did provide them with lots of material to take away, and we want to make sure that that's our overall conclusion. We had listened to the small-group discussions and read the written comments and the papers that were commissioned for this Forum; we all concluded that a little listening and reading between the lines were needed to pull out some of the themes. So my part of the summary is to try to identify a few of the things that we found between the lines, maybe not mentioned as prominently or explicitly as we had expected.

For example, on my first reading, I thought Richard Tapia's article was brilliant, really very well reasoned, very convincing, persuasive, certainly on an important topic, and utterly unresponsive to the topic he was asked to address. From discussions with other people here, I learned that many had similar reactions. But by the end of yesterday, after listening to the discussions, I concluded that, not only was Tapia's paper responsive, it was perhaps the most important take-home message from the Forum and very relevant to assessment. College admissions, the use of tests and grades, and the grading system itself are all

[1] During 1998-99, the College Level One team of the National Institute for Science Education (NISE) will be organized into an "Institute" that focuses on assessment in postsecondary SMET education. See *http://www.wcer.wisc.edu/nise/cl1*

examples of things that clearly have huge impacts on everything we do in higher education. And these processes and systems produce outcomes that can be checked against goals. But if we aren't clear on what the goals are, then we shouldn't be surprised if we look at the outcomes and don't like what we see. The public at large, the courts, and many faculty believe that grades and test scores can be used to generate a fair, accurate, and objective one-dimensional scale of admissibility. To the extent that we fail to focus and agree on what we are trying to predict at the time of admission and fail to assess what we are actually predicting with the present system, we're misleading everyone, including ourselves. This issue may be the single most significant one in the whole discussion of educational reform and assessment.

Let me mention a few other things that we found between the lines. Curriculum was mentioned in the title of the very first panel, but we thought it wasn't sufficiently in evidence in either the panel discussion or the audience participation and discussion that followed. This is just one aspect of the larger issue of alignment and that did emerge, as Cora told you, in many of the discussions and the written comments. But we felt curriculum is important enough to be highlighted. Curriculum in particular is not just a collection of courses, even if those courses are selected from a constrained list or menu. Improving courses and improving pedagogy in individual courses, one course at a time, may have some value, but it won't get us very far in the overall reform effort. Someone talked about cattle-call courses; those required courses taken by large numbers of students who don't really want to be there. As an engineering faculty member, I can tell you that some of them shouldn't be required to be there. We have not done a very good job as engineering faculty in deciding exactly which courses are really essential prerequisites for many of the engineering majors. My own department of electrical and computer engineering has multiple subpaths through it, and it's almost inconceivable that the software engineers, just to pick one example, genuinely need chemistry. It may be nice for them to learn chemistry; I took a lot of chemistry myself as a student, and it was a valuable experience. But many of the students who choose software engineering as their major do so because they have discovered they have no aptitude for and certainly no interest in chemistry. It is going to be very difficult to motivate them strongly to take a course that they know, better than we, they aren't going to use subsequently. So thinking about the curriculum involves a lot more than just learning how to teach chemistry better so that these students will benefit more from it. You can all make up your own examples. I am sure we've all had experiences with courses that were taught out of phase, or that students were encouraged to take in an inappropriate order. The total quality management experts and consultants who are very much in evidence these days like to make the distinction between doing things right, on the one hand (e.g., improving courses), and doing the right things on the other hand (e.g., deciding which courses and which topics are really important).

The third major thing that we found between the lines is the whole issue of scalability. Cora touched on that briefly—and maybe this is just a thick-headed, bean-counting, Philistine provost's perspective here—but it is important to keep in mind that we are educating students at the postsecondary level for about the same per student cost as day care. As someone who recently had two children in day care, I can tell you that $7500 a year per child for day care is not that different from the amount that most public universities spend per student per year. No one is going to give us lots more money, either to do things better, or to do assessment for any reason whatsoever, whether it's to improve our internal processes or to provide the required external accountability. I have no doubt that it is possible to design educational, pedagogical,

and other improvements that are wonderfully effective but so expensive that we can't afford to scale them up and do them across the curriculum or for all students. A very good example, from a completely unrelated field, is automobile safety. If you've ever watched the Indianapolis 500 or some other race, you know it's possible to have a car that goes 150 or more miles per hour, crashes into something, tumbles end over end for a hundred yards, and bursts into flames, and then to have the driver step out of the car and walk away. It is possible. But you and I cannot afford to drive cars that are designed to that standard. We need to keep affordable standards in mind both for our course improvements and for any assessment that we find important to add on. Assessment can't add much time or cost if it is to be scalable. Again we believe that affordability is a very important issue for the CL-1 team members to keep in mind as they study assessment over the next year.

Finally, it is important to spend some time thinking about developing a rhetoric for success. Everyone here believes firmly in the need for educational reform, the importance of changes of various sorts, although we don't spend much time among ourselves here talking about exactly what change. As you lead the charge for educational reform, *change* should not be the word on the banner that you carry. Manuel Gómez said this morning that faculty want to do a good job, and by and large most of them think they are doing a good job. A frontal assault with the banner of change is very unlikely to get people to confront the data and make the changes that you are looking for. When I was dean of the graduate school, we were very concerned about the time to Ph.D. degree; this was in the early 1990s, roughly the time of the Bowen and Rudenstine book on that topic. We quietly did an internal study in the graduate school and gathered all of the facts on time-to-degree by major across the campus. We also asked every department chair how long, on average, it takes students in the department to get a Ph.D. degree. Almost everyone answered, and the overwhelming majority of the answers were in the four- to five-year range. In fact, in many departments the average time to Ph.D. was in the eight- to ten-year range, and in lots of others the mean was in the six- to eight-year range. Very few departments, mostly in the sciences, had times that were anywhere near four or five years for either median or mean. To do something about this discrepancy, we printed out for each department a list of all students who were still enrolled, and had been certified as making satisfactory progress to their degree, and were five or more years past the preliminary exam. Since it usually takes three to five years to get to the prelim, those students had been around a long time. We sorted each list chronologically so that the first name on the list was that of the student who had been there longest. One department's list was two pages long. The chair of that department wrote me back thanking me profusely for the wonderful and useful list that I had sent him. He said that the department hadn't even known a couple of the students were still enrolled! Subsequently that department reexamined what was going on and has almost completely flushed out, in most cases *with* degrees, the students who had been around inappropriately long and were not making progress. That department simply needed to know the facts. As Manuel said this morning, getting the facts out is the most important step in stimulating change.

The rhetoric associated with assessment also needs some attention. Cliff Adelman pointed out that content experts are not assessment experts, and very often the reverse is also true. While it is natural for assessment experts to develop their own vocabulary for talking among themselves, using a new vocabulary has the unfortunate consequence of creating an us/them or insider/outsider dichotomy. Most people believe they know what the word *assessment* means. Certainly most people working in higher education believe they know what it

means, and they think they are doing it already. We don't believe it will be productive to lead them to conclude that assessment is some arcane discipline that has nothing to do with anything they have ever done before or anything they've ever thought of as being assessment. Massive amounts of new vocabulary will perpetuate the us/them dichotomy and virtually guarantee that assessment will not be easily or widely adopted.

Cliff also pointed out that there's a danger in getting feedback from the assessment that is so negative that it is simply suppressed or ignored. I think that this danger is probably only a short-lived one if assessment is really an integral part of the educational process. It is not possible for a system in equilibrium to get very far away from its set point if there is continuous feedback and correction. So if you set new goals, an assessment showing that you are very far from where you want to be can initially seem very negative. But if you continue the assessments, results will show that you're steadily nearing your goal and will identify small areas for further improvement. By and large, such feedback will be welcomed and heeded. If you emphasize the increasingly welcomed feedback that comes from assessment, you'll have a much easier time of getting it adopted and accepted. The whole point of assessment is to make surprisingly *large* negative feedback impossible.

In the session this morning it was pointed out that accreditation reports often just go on a shelf until the next visit of the accrediting organization. At UW-Madison, we've been trying to make sure that accreditation visits are a useful exercise. One of the requirements of the regional accrediting associations is that institutions do a massive self-study involving a very large number of people from across the campus. We identify representatives of all the governance groups to do the study, so the document itself and anything that comes from it automatically has a great deal of buy-in and legitimacy. After our last North Central accreditation, we asked that the governance group representatives remain convened after the accreditation visit was over and write a brief summary document that would be a road map for the ten years until we would be accredited again. That document has been extraordinarily valuable to us over the last ten years. At about the halfway point, five or six years into the ten-year period, we went through a formal exercise of rewriting the document and re-expressing things in what seemed to be more up-to-date language. We treated the product as a vision document for the entire campus. Although it isn't a strategic plan for the campus, it lays out some principles and things that we believe everyone is or should be working toward. All of the deans were told to take that back to their colleges and develop a strategic plan that was consistent with the overall campus vision document. Each college strategic plan was then given to each department to develop a strategic plan that was consistent with the college plan. In this way, we have developed a nested set of plans that are more or less in alignment. In view of this morning's discussion about the danger of ignoring accreditation visits for ten years, I suggest our generic model as one that other institutions might adopt.

Analysis of Participant Theories of Change

Elaine Seymour

Embedded in the papers, panel discussions and think piece essays were a set of diverse theories about how the improvement of SMET higher education could best be secured, the conditions that enable or constrain success, and the integration of assessment into change strategies used at all levels—classroom to nationwide. The theories reflect differences in the ways participants defined situations and agreed or disagreed about the best courses of action. These are not "armchair" theories: they reflect the ways in which a group of people are actually approaching the task—a group whose collective efforts may affect the direction of change in SMET higher education.

Diverse Theories of Change

Some participants' theories are diagnostic or explanatory: they focus on the sources of issues and the ways in which these issues may be addressed. Some are theories of action: they advocate and explain the rationale for particular strategies. Threaded through these theories are participants' views about the roles to be played by assessment and evaluation. Both types of theories have powerful practical significance because they arise from the hands-on experience of people who represent an important subset of the educators, researchers, administrators, and policy specialists currently engaged in the nationwide movement for improved quality in SMET higher education.

Change Is Driven by Shifts in Values

Forum discussions highlighted two shifts in educational priorities that are seen as driving change in SMET higher education. The first is a shift in the emphasis in classroom work from teaching to learning, while the second entails a shift from SMET education focused largely on science-for-the-few to the nationwide goal of science-for-all. These two shifts are not disconnected: a focus on learning in SMET classrooms is critical if the vision of science-for-all is to become a reality. Moreover, new assessment criteria and methods play an important role in each.

Forum participants indicated that the shift at the course level from teaching to learning includes:

- rediscovery of the value as professional activities of teaching and of education scholarship and research;
- refocusing classroom practice on gains in student understanding, reasoning, application, and learning retention;
- alignment of course goals and assessments to provide instructors with feedback on the efficacy of their work and to engage students in their own learning;
- rethinking professional relationships with colleagues in K–12 education, science education, assessment and evaluation research, other SMET disciplines, and the social sciences and humanities; and
- rethinking professional education and development, including re-education for mid-career faculty, and attention to the types and quality of primary professional education offered to preservice K–12 teachers, graduate students, postdoctoral fellows, and entering faculty.

Developing panelist David Porter's point that every system is perfectly designed to yield the results observed, Daryl Chubin (discussant) explained that the change from teaching to learning requires a shift from "valuing what we measure to measuring what we value." It begins with rethinking what constitutes "success." It requires that we place primary emphasis on learning, and not only assess gains in learning outcomes, but

also improvements in the process of learning. As panelist Diane Ebert-May explained, we must learn to measure "active knowledge (understanding, reasoning, and utilization) rather than discrete, isolated bits of inert knowledge."

The second shift in educational priorities is from SMET education focused largely on science-for-the-few to the nationwide goal of science-for-all. This shift reflects, as panelist Eric Mazur observed, recognition that competitive global market realities now require all educated citizens to be science and math literate in order to function effectively. To undertake this shift, campuses must cast a critical eye on their entry-level classes, where it is standard practice to use assessment to locate "the few" by weeding out the rest. This practice, which actively hinders achievement of the emerging science-for-all goal, has serious societal and moral implications. Once having acknowledged this problem, campuses must then learn which teaching strategies most effectively provide an adequate science and mathematics education to all college students—and then broadly use these strategies. As panelist Richard Tapia reminded us, SMET faculty have the collective power and opportunity to change the conditions that have created a permanent and growing under class in our society, one result of which is limited job options due to lack of scientific, mathematical, or computing skills.

Departmental Values Are Key to Educational Improvements

The Panel Two participants, and many of the think piece writers, pointed to the pivotal role of departmental values and practices in enabling or obstructing efforts to make these shifts. What we value in teaching is operationally defined and evaluated at the department level.

Panel facilitator Elaine Seymour pointed out that, traditionally, the effectiveness of teaching has been evaluated informally by peers and formally by institutional or departmental end-of-semester student classroom evaluation instruments. The criteria faculty use to judge each others' teaching are made explicit in collegial arguments about curriculum change and in course evaluation criteria (e.g., good organization, clear and lively delivery, good "coverage" of accepted topics, strict adherence to grading standards, and careful observance of formal duties). These criteria reflect a view that the teacher's primary duty is to pass on the standards and the canon of the discipline. To establish and guard these discipline standards is viewed as a primary responsibility of the department.

When confronted with colleagues who have shifted to a focus on learning and/or learning-for-all, faculty tend to assume that the new foci threaten departmental effectiveness in preserving and transmitting the canon. Panel facilitator Brock Spencer illustrated this assumption in action when describing his colleagues' concern that his "lectureless" approach to teaching undergraduate chemistry (in which much of the learning is done in small groups) "lacked rigor." In particular, colleagues cited the dropout rate in his class (which was almost nil) as evidence of insufficient rigor, by comparing it with that in traditional sections, where it often approached 30%. As this example illustrates, the departmental sense of responsibility to the canon is very strong. It is so strong that faculty must experience, as panelist David Porter put it, "a paradigm shift" before they can see that the adoption of these new foci do not threaten, but actually support, the departmental mandate to pass on the standards of the discipline.

Brian Coppola (panelist) argued that it is for this reason that merely providing departmental colleagues with findings from educational experiments "is not an effective agent of cultural change." "Information," he explained, "is received within a cultural context that uses its own criteria to decide what, if any, use to make of that information." To take note of the data and use it to inform a plan also requires a value

shift. The best way to foster this value shift, he argued, is to instill in a new generation of science faculty a value system that embraces education as scholarship.

Jack Bristol (panelist) thought this process was already underway. He argued that the market was on the side of the change makers: the shrinking availability of assistant professor positions fuels interest in education among graduate students and postgraduate job seekers. Several presenters and think piece writers advocated hiring some new faculty in each department specifically because of their educational specialties and educational research skills. The education of graduate student teaching assistants in the teaching methods required for what Cora Marrett (closing speaker) called "productive education" thus becomes a key strategy in faculty efforts to scale up learning-focused approaches to curriculum and pedagogy.

Rebalancing the Departmental Rewards System to Reflect Respect for Teaching and Educational Scholarship

There was a strong consensus among both panelists and think piece writers that the faculty rewards system is the primary barrier to efforts to shift the emphasis from teaching to learning. One of the strongest single messages sent forth from this Forum was that change cannot occur unless departments alter the criteria used to distribute rewards. Participants noted two primary consequences of failing to find effective means to address this barrier to change: (1) loss to the reform effort of the next wave of young faculty—through pedagogical knowledge not gained and socialization into prevailing departmental attitudes and practices while waiting for the safety of tenure; and (2) casualties among the current generation of creative teachers. In light of these negative consequences, it was not surprising that no panelist was willing to encourage untenured faculty to devote significant time to teaching. As Richard Tapia observed, "You can't change the system if you're not there."

Participants proposed several ideas for how changes in the criteria for departmental rewards—tenure, promotion, resources, time, and opportunities for professional development—could be made. For example, panel discussant Norman Fortenberry proposed that departments extend tenure and promotion guidelines and criteria to include "peer-recognized intellectual work, appropriately disseminated" as a way to align teaching scholarship to existing departmental values. (He reported that this had been accomplished at Oregon State University.) Changes such as this not only would promote improvements in the quality of higher education teaching, but also protect the current generation of innovators, promote institutionalization of their work, and allow younger, untenured faculty to choose classroom scholarship as a professional activity. Other participants suggested that, to implement reward criteria that more truly value teaching, it will be necessary to measure teaching competence in terms of indicators of student learning gains and to recognize contributions to peer-reviewed education scholarship. Participants stressed that these and any other strategies proposed for making changes in the reward system must balance respect for pedagogy with respect for disciplinary research.

The most common strategy that was proposed for establishing departmental rewards that balance teaching and research entailed the exercise of leadership—at both local and national levels. In particular, panelist Sheri Sheppard cited accreditation agencies, public and private funders, and national testing services as external bodies with the power to create a shift in values and faculty priorities. She was supported by other panelists and writers in expecting college presidents, provosts, and deans to take a lead. At the same time, these contributors also acknowledged the power of departments to effectively resist the goals and plans of

institutional leaders and even the exhortations of national leaders.

Evidence Is a Necessary but Not Sufficient Condition for Reform

Many of the think piece writers shared the view of James Highsmith (participant) that "the most useful thing we could provide for institutional change would be clear, concise, and valid data that new approaches to education are as effective as, or more effective than, traditional approaches." This is necessary because, as Ken Verosub (participant) explained, "few of the laggards and the cynics will accept the need for change until they are shown that their preconceptions are wrong and that other models do work." As Cora Marrett pointed out, "Evidence counts. It is not enough to claim that productive education can occur; clear and convincing evidence matters." Failure to supply evidence was seen by many contributors as an absolute barrier to change. However, for a number of reasons, providing this evidence is not a simple matter. Moreover, the resources expended in providing proof must be understood as merely a necessary, and not a sufficient, condition for persuading one's colleagues to change.

One reason it is not simple to collect proof that new approaches are effective is that one must wait until reformed courses have matured before it is appropriate to collect data on their outcomes. As Eric Mazur and others experienced, "The initial effect of any change is not an improvement, but a period of problems, adjustments, mismatch, and, to some extent, frustration." To collect outcome data while faculty and students are still adjusting to the new methods may be to collect misinformation. A second reason it is not simple to collect proof is that faculty expect data that compare traditional and reformed courses, but it often is not meaningful to use the same assessment tools to gather data on traditional and reformed courses. As Mazur argued, changed instructional methods require a matched change in assessment methods. The change in assessment, in turn, "means giving up any meaningful correlation with previous assessments. As long as administrators and faculty do not realize that this poor correlation is an unavoidable consequence of change, it will be impossible to move forward."

Questions of difficulty aside, a number of participants pointed out that providing evidence of the effectiveness of reformed courses is rarely sufficient, in and of itself, to convince others to change. Brock Spencer illustrated this point from his own experience: although his chemistry colleagues acknowledge that the students from his lectureless class do at least as well in subsequent classes as those who elect more traditional sections, they are still not convinced of its "rigor."

Indeed, some participants questioned why change makers are required to expend their resources developing evidence of rigor when it is not at all clear that the assessment practices used in traditional courses provide effective evidence of learning. As Earl Peace (participant) wrote, "There is no evidence that the traditional courses are as 'successful' as some cynics would claim. [People] hold reformed courses to a higher and different standard of proof." The identification of this double standard led Coppola to observe that faculty have long assumed that the traditional array of tests, examinations, papers, reports, and projects adequately measure what students have learned in class and lab. Only recently have more than a few thought to examine critically how well these tools assess the degree to which students either learned or retained what they had been "taught."

While many participants were concerned about the above-noted problems implicit in the search for data that prove the effectiveness of reformed methods, they nonetheless believed that certain types of assessment data are important to collect. Adelman and Gómez, of the third panel, discussed the need to gather institutional

data—such as enrollment and graduation rate—and then effectively communicate findings that indicate success at course, department, and institutional levels. They emphasized the need to "translate" raw data into information that faculty and administrators can understand and value, and to use presentations designed to "get the point across" to audiences with different priorities and levels of sophistication in the interpretation of data. For example, members of both Panel One and Panel Three suggested that one effective way to communicate with faculty is to focus on indicators of outcomes that faculty of all pedagogical persuasions are likely to value—such as the understanding and skills students should carry with them from one class to another or into the real world. In seeing data that relate to their own standards of professional practice, faculty are more likely to feel obligated to address (through new practices) the inadequacies highlighted by assessment findings.

In spite of the difficulties and frustrations implicit in providing evidence that new approaches are effective, course innovators and the professional evaluators with whom they work are engaged in a shared search for forms of assessment that better serve course learning goals. The Panel One presenters and their discussants offered examples of such assessments from their own professional experience and pointed to others. The need for such information is clearly great: many think piece writers voiced their wish for easier access to the body of existing work on new assessment methods.

Alignment Is Required at All Levels for Effective System Change

The need for alignment was a recurring theme throughout the papers, panel discussions and think pieces. Panel One members focused on the steps needed at the classroom level to make the curriculum more meaningful to students. Faculty must articulate their learning goals and align their teaching and assessment strategies with those goals. Panelists advocated making students aware of these connections by "sign posting" and by making students aware of their own learning processes. These activities help both students and instructors understand and benefit from the alignment between course goals and strategies.

It is also alignment that makes a decentralized system coherent. In contrast to systems in many countries, where the linkages among system components are formally planned and orchestrated at regional or national levels, the U.S. system is built on a convergent but local and independent basis. Thus, we cannot take alignment for granted. Our system is also more difficult to change, because, as the discussion of the third panel reminded us, we must first identify the most powerful elements in the system and then figure out how to leverage them to create change in order to make progress. The participants' concern with alignment reflects their understanding that the attempt to alter single elements in a complex social system will not be effective. Each element must be aligned with the others for system change to prevail.

Discussions across the Forum highlighted five main alignment issues, each of which participants viewed as critical components of nationwide reform.

- Alignment of classroom assessment practices with student learning goals (a major theme in the first panel session)
- Alignment of teaching endeavors across departments with the overall teaching mission of the institution, and the use of classroom assessments as essential building blocks for the evaluation of institutions overall
- Alignment of curriculum developments in SMET higher education with developments in K–12 reforms (Chubin argued that "we should think harder, and more seamlessly, about the education continuum as K–14 or K–16")
- Alignment of the activities of SMET faculty classroom reforms with the

knowledge and skills of colleagues working in the disciplines of education, assessment, and evaluation
- Alignment of data collection practices such that national information can better inform evaluation practices at the institutional level, and classroom assessment can meet both departmental and institutional needs for evidence of efficacy, as well as the needs of the teacher and learners for feedback.

Bottom-Up and Top-Down Theories of Change

Grass-roots Action. At least three of the panelists (David Porter, Eric Mazur, and Eileen Lewis) and many of the think piece writers supported the theory that reform could be generated through grass-roots action. Porter used the metaphor of reform ideas as a virus that other people could catch. Lewis spoke of change that begins at the fringes, and others wrote of "community-grown" efforts from which change ripples out and upwards. These speakers added caveats that departments and institutions need to recognize and reward the individuals and small faculty groups undertaking reform if the best practices are to grow and thrive in the longer term.

A Network Theory of Change. A second, more structured, version of the bottom-up theory proposed that change can be built from small local beginnings, first by provoking and maintaining conversations that lead to local collaboration, then by making connections with collaborators on the same or other campuses. Many of the Forum participants had been engaged in, and sustained by, such conversational networks. These writers, particularly those more isolated faculty, viewed the process of working out the details of new professional practice in the classroom in the course of such conversations as a source of intellectual stimulation, new learning, and peer review. It sustains the innovators, especially those in situations of indifference and risk. Although no one provided an example of networks of such collaborations that built into a critical mass, many hoped for such an outcome. Several writers pointed to the desirability of collaborations across disciplines, but only a few offered good suggestions about how these might be evoked or sustained. Sheri Sheppard was among those who perceived that cross-departmental networks will be needed to make effective educational use of the rapidly emerging computer-based communications technologies. She also noted that many major questions about the use of these new technologies—the impact on learning of different types, their relative cost-effectiveness, and their likely impact on assessment practices or the distribution of faculty rewards—remain to be answered.

A number of contributors pointed to the role that national groups could play in supporting and extending conversational networks and initiatives. Several viewed the NSF and NISE as examples of bodies with the resources and scope to enable such conversations. Participants suggested that these agencies should provide more workshops and working meetings/ conferences, travel money for meetings, funding for Web site maintenance, publicity, and recognition of individual or group achievements.

Value-driven Institutional Leadership. Members of all three panels and many think piece writers countered that change within departments can not occur without institutional leadership. As Joel Mintzes (participant) explained, bottom-up approaches were ultimately fated to fail unless "catalyzed and sustained through intervention of enlightened and benevolent administrators," who support change by the strategic distribution of resources, funding, and recognition. Manuel Gómez, among others, asserted that system change requires unequivocal, high-level commitment to the values of science-for-all and learning-centered teaching. Administrators must provide a coherent, institutionwide system of professional rewards commensurate with student learning gains and of assessment

practices developed in negotiation with those whose work is to be judged by them. As Alberta Lipson (participant) observed, if these things "are valued at the top, resources will be found and faculty will be committed." Administrators need to plan for change at multiple levels, be ready to spend money to make change work over the long haul, and work for buy-in at the critical departmental level. Writers also exhorted administrators to identify, support, and recruit into their plans "natural" faculty leaders and to reinforce creative collaborations. In this regard, some cited the impact on the departmental rewards and value system that could be made when respected members of the senior faculty took an active, public interest in pedagogical renewal.

The Blueprint Model: Progress Depends on the Accessibility of Proven Models, Practices, and Assessment Tools

The Panel One speakers, and many of the think piece writers, agreed that, regardless of where change begins and how it is sustained, the efforts of faculty interested in trying new methods will be greatly enhanced if they have better access to information about well-tested teaching and assessment methods. They also saw promotion of models of change from different institutional contexts as critical. Time, effort, and resources should not be wasted on strategies that did not work well in other comparable settings.

A large number of Forum participants expressed a strong demand for information, examples, assessment tools and methods, and access to pedagogical and assessment expertise. They sought ready-to-use teaching and assessment materials, requested digests of techniques, and evidence of their efficacy—including reports of what has not worked well. As Senta Raizen (NISE team leader) pointed out, because information about what does not work rarely appears in peer-reviewed journals, reformers may get the false impression that all evaluations find positive outcomes. With this point in mind, several writers and speakers emphasized the need for a venue for reports of failed experiments. Participants expressed a need for syntheses, based on published and unpublished reports, of what we have learned about methods that do and do not work, and why. Many writers asked national agencies to provide leadership in developing and maintaining permanent electronic repositories containing curricula, assessment tools, teaching methods, syntheses of the experiences of reformers, and other resources. Some suggested that the repositories be indexed, reviewed, and continuously updated; provide guidance on how to use innovative materials; and provide information about how to locate disciplinary colleagues with similar goals and people with relevant expertise in other disciplines, such as science educators or social scientists with assessment or evaluation skills. In his opening remarks, Luther Williams described the NSF's first steps in developing a national SMET digital library and the agency's awareness of the need "to enhance the coherence between faculty development needs and curriculum development."

Participants also encouraged the funders of educational initiatives to make their program guidelines for evaluation and dissemination more specific and demanding. They should require both the provision of formative feedback to participants and evidence of progress in using the evaluation design to build ongoing self- and unit-monitoring practices.

Faculty also sought workshops and other professional development opportunities to build their own teaching and assessment skills, and those of junior peers and graduate students. These strategies again imply change through leadership—by policies, action, and sustained funding at the national level. In this regard, Norman Fortenberry described the NSF's Online Evaluation Resource Library and the proposed SMET National Digital Libraries Initiative, and Arthur Ellis (NISE team leader) described the Field-tested Learning Assessment Guide

(FLAG), developed jointly by NISE and the New Traditions Chemistry Consortia.

Throughout the discussions of resources needed to improve the quality of undergraduate SMET education, participants made intermittent reference to under use of the collective expertise in curriculum development and in pedagogical and learning assessment techniques that resides in colleges of education. Many four-year campuses have colleges of education, making access to such expertise and informational resources readily available to faculty and administrators. A number of panelists and writers noted that a long-standing cultural difference between SMET and education faculty continues to inhibit collaborative relationships among these faculty, despite the NSF's current emphasis on enhanced preparation for K–12 science and mathematics teaching.

Change by Leverage from External Agencies

Sheppard, Bristol, and others cited powerful national bodies—the accreditation agencies, the national testing services, the national scientific societies, and the National Academy of Sciences—as having individual and collective power to leverage change within institutions and departments. However, only recently did many see this leverage being used to promote education reform. Members of the third panel and contributors from the floor during this session portrayed the accreditation agencies as a historically conservative influence. They described the accreditation agencies' limited use of assessment techniques as contributing significantly to the predominant faculty view of evaluation as irrelevant and burdensome. Panel Three also discussed a recent trend to link accreditation activities with education reform efforts. For example, panelists explained that the Accreditation Board for Engineering and Technology is in the process of fostering a national change in the engineering curriculum and in faculty attitudes toward engineering as education. Similarly, Jack Bristol credited the changes he has witnessed over the last two decades at the University of Texas-El Paso as driven mainly by the demand for better evaluation coming from external sources such as the NSF. He saw this as an essential impetus for change efforts within his institution.

Other national agencies (particularly the scientific societies) drew a smaller share of Forum attention. Several think piece writers proposed more national discussion of their influence—especially at the departmental level—in setting normative and structural limits on SMET curriculum development, classroom assessment practices, and cross-disciplinary collaboration.

Participants expressed mixed feelings about one long-standing model of reform that uses external agency leverage, namely, grants-driven reform. By this strategy, groups of innovators are funded to develop and test an educational initiative in the hope that it will take root in the host institutions beyond the end of funding. This pump-priming strategy for change has been handsomely underwritten by the NSF and many private foundations over at least two decades. Based on their own experiences as current or former recipients of such awards, many Forum participants cited the positive role of outside funding in testing educational experiments and "catalyzing change." At the same time, the Panel Two participants and some think piece writers pointed to unfortunate side effects of a system of funding for educational experimentation that parallels competitive research funding. It focuses collegial esteem on the size and renewal of awards for educational experiments, rather than on their evaluated results and their longer term implications. This focus, in turn, reduces the chances for changing departmental values and, thus, for the survival of the initiative—unless host institutions undertake responsibility for the initiative. To overcome this problem, participants suggested more funding of programs that adapt models with an established record and more funding for

adopter/adapter groups. A second alternative, outlined by Norman Fortenberry, is to provide more funding for higher education research within the SMET disciplines and for curriculum development projects that build on such research. A third alternative is to fund more organization-wide projects where institutional buy-in (including matching funds and resources) is guaranteed from the outset. Chubin expressed the hope that "federal dollars can foster cultural change and build human capital, especially at institutions that can assist faculty in becoming true scholars: custom-oriented assessors, facilitators, and mentors to the next generation of scientists, engineers, and citizens."

Conclusion

Considering the range of theories of change that were explicitly or implicitly presented in the documents produced for and during this national meeting, we conclude that the community represented at the Forum has now begun to understand that assessment is critical to success in reforming undergraduate SMET education. Forum participants were considering—or had already begun to experiment with—assessment practices that are aligned with new goals for student learning. They were also seeking resources to help them with their experiments. Upon learning that the assessment efforts of colleagues who already have ventured down this path are neither well documented nor widely available, Forum participants called for commissioning, collection, review, codification, and dissemination of such resources.

In their discussions of the kinds of learning environments that are needed if emerging assessment practices are to prevail, the participants returned repeatedly to the need for systemic change. They cited the need for change not only in the values and practices of individual instructors, but also in the values, policies and practices of departments and whole institutions. They also urged that changes in individual courses, departments and institutions be aligned—across academic disciplines and departments, across postsecondary institutions, and with the K-12 system. Thus, in addition to the call for resources to enable the development of good individual assessment practices focused on student learning, the Forum participants called for fundamental reform of the undergraduate SMET education enterprise.

As a final word on this Forum, we note that the issues raised about postsecondary SMET education assessment also pose new questions for the research and evaluation community and for education professionals. These include:

- How shall we measure faculty efficacy both in their teaching role and as educational scholars?
- How can we best reward these professionals?
- How will faculty develop the skills they need to do either kind of activity well?
- How can we raise the quality of the current and future mathematics and science K-16 teaching force on which all else depends?

Appendix A
Panelist Papers

Panel 1:
Assessment of Teaching, Learning, and Curriculum Change in SMET Classrooms

Assessment as a Learning Process: What Evidence Will We Accept That Students Have Learned?

Diane Ebert-May
Science and Mathematics Learning Center and
Department of Biological Sciences
Northern Arizona University

As scientists, we spend a considerable portion of our professional time gathering information and making decisions based on that information. We read journals, attend conferences, and collaborate with our peers to obtain information that will improve our strategies for conducting research. We decide whether or not the strategies are appropriate for use in our research. As teachers, we also gather information, in this case, about our students. Based on that information, we make decisions about student learning and our teaching practice. We engage in the process of gathering and interpreting information and making decisions based on that information—we assess (Champagne & Ebert-May, unpublished data).

The type of information we gather about our students depends on the evidence we will accept that the students have learned what we wanted them to learn. We must have confidence in the quality of the information to justify our subsequent decisions about teaching. Major changes in assessment based on measurement theory and practice have catalyzed the development of new methods of data collection along with new ways of judging data quality. If, indeed, learning science should be an active process (Ebert-May, Brewer, & Allred, 1997) then assessment should measure active knowledge (understanding, reasoning, and utilization) rather than discrete, isolated bits of inert knowledge. The new view is that "assessment and learning are two sides of the same coin" (National Research Council, 1996, p. 76). The methods used to collect educational data define in measurable terms what we should teach and what students should learn.

Hodson (1992) described good assessment procedures as fulfilling at least four functions. First, a summative function: assessment should provide some description of students' levels of attainment in all components of the course. Second, a formative function: assessment should provide diagnostic feedback to the instructor and students throughout the course about the students' strengths and weaknesses, understandings and misconceptions to more effectively plan further learning by each student. Third, an evaluative function: assessment should provide instructors feedback about the effectiveness of the curriculum to assist ongoing decisions and planning about curriculum. Finally, an educative function: assessment should engage students in interesting, challenging,

and significant experiences aimed at helping them develop further understanding and reasoning skills. When used as an educative tool, assessment becomes part of learning. Students actively participate in assessment and, by doing so, move toward taking responsibility for learning and advancing themselves as independent, life-long learners (Angelo & Cross, 1993).

The reform of assessment in this country is led by the K–12 segment of the educational continuum (NRC, 1996), and there is a critical need for higher education to join in this process. If we want to know how faculty can understand best what undergraduates are learning, retaining, and using in future contexts, we need to approach assessment with the same level of knowledge, rigor, and confidence as we do the collection and interpretation of our scientific data. So too, approaches to implementing new forms of assessment should be comparable to utilizing new laboratory techniques. A scientist would seldom adopt a new laboratory technique without considering the purpose of the technique, the influence of the technique in the context of the experimental design, and the potential consequences of the technique in terms of results (Champagne & Ebert-May, unpublished data). Change in assessment practices requires similar thought and examination.

Aligning Assessment with Learning: A Case

In our introductory biology course designed to develop biological literacy for all students, one goal was for students to effectively communicate to peers and others an understanding of, and links among, biological principles and concepts (Ebert-May et al., 1997). What evidence did we accept that students adequately communicated that understanding? We defined the tasks and performance standards appropriate for students in an introductory nonmajor biology course guided by these questions: What type of written and oral communication assessment projects or tasks are appropriate? What are the biological principles that students must understand to communicate the ideas? What criteria would we use so both the students and instructors could differentiate levels of performance in both written and oral communication about the biological principles?

While our choice of assessment techniques to collect evidence about students' accomplishment of this goal was multifaceted, the underlying principle driving the choice of assessment was straightforward—the assessments we choose must enable students to communicate in both written and oral format their understandings of and links among biological principles. Simply stated, the choice of assessment form must be consistent with the student goal and what the instructor intends to infer from the data. The strategies we employed to address this goal included short writing samples, essay questions, quizzes, concept maps, self-evaluation, peer review of papers, class discussions, public hearings (Brewer & Ebert-May, in press) and, in the laboratory, oral presentations of research proposals, research papers, and poster displays.

For each of these assessment strategies we developed a scoring rubric. Scoring rubrics are the specification of the knowledge and ability components or the product characteristics as well as the point value assigned (Champagne, unpublished data). Rubrics define the performance standards for a population of students based on a desired student outcome. For example, students addressed questions with short answers in class to communicate their understanding of and links among biological principles. For this task we defined the criteria for various levels of achievement (Table 1). Students were given this scoring rubric at the beginning of the course and were encouraged to use it as they wrote. We also solicited feedback from students regarding their understanding of and input to the criteria on this rubric. For other assessment strategies, students developed rubrics with their

Table 1. Scoring rubric for short writing samples completed by students during class. Students were assessed on both their general approach and comprehension.

Level of Achievement	General Approach	Comprehension
Exemplary (5 pts)	• Addresses the question • States a relevant, justifiable answer • Presents arguments in a logical order • Uses acceptable style and grammar (no errors)	• Demonstrates a clear and complete understanding of the question • Backs conclusions with data and warrants • Uses 2 or more ideas, examples and/or arguments that support the answer
Adequate (4 pts)	• Does not address the question explicitly, although does so tangentially • States a relevant and justifiable answer • Presents arguments in a logical order • Uses acceptable style and grammar (one error)	• Demonstrates adequate understanding of question, but does not back conclusions with warrants and data • Uses only one idea to support the answer. • Less thorough than above
Needs Improvement (3 pts)	• Does not address the question • States no relevant answers • Indicates misconceptions • Is not clearly or logically organized • Fails to use acceptable style and grammar (two or more errors)	• Does not demonstrate understanding of the question • Does not provide evidence to support the answer to the question
No Answer (0 pts)		

instructor. This approach has the potential to increase learning by engaging students in performance assessment.

Our choice of assessment depended on the desired student outcome. For example, to gather formative feedback we used concept maps (Novak & Gowin, 1984) and writing samples to assess students' understanding of the links among biological principles. So for a quiz, students were provided a list of concepts (5-6) and asked to develop a concept map. Alternatively, for homework students were asked to identify the concepts for a set of readings and build a concept map. To prepare students for this type of assessment, we modeled the use of concept maps in class. Our intention was to use multiple forms of assessment to provide the kind of evidence we needed to make decisions about teaching. Figure 1 shows a concept map about innovation in teaching and learning that represents meaningful relationships between concepts in the form of propositions. The concepts are arranged hierarchically and provide a visual map to benefit both faculty and students.

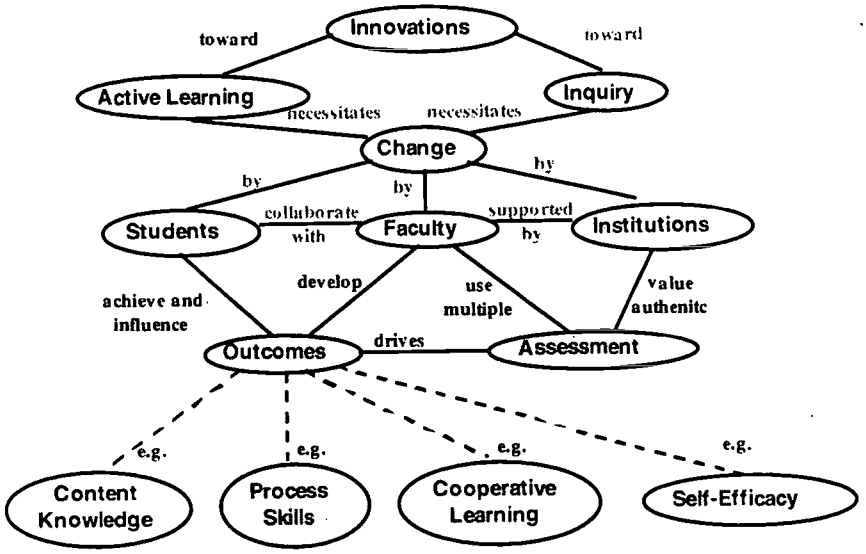


Figure 1. Concept map representing the relationships between concepts about innovation in teaching. The concepts are arranged hierarchically and the connections between concepts form a logical proposition.

The nature of evidence faculty are willing to accept about student learning and the way they go about collecting and interpreting the evidence provide the best guide to inform their decisions about teaching. Importantly, the student goals and assessment must be based on acceptable, well-defined criteria that faculty will accept as evidence of student achievement. If faculty value students' abilities to solve interdisciplinary problems in biology, assessments must include opportunities for students to demonstrate their ability to solve problems. This criterion does not minimize the importance of understanding content. Rather, the issue is how to assess students' knowledge in ways that demonstrate their in-depth, long-term understanding of content.

Self-evaluation: A Way to Reduce the Risks of Innovation

During this session, several interrelated themes focus on the nature of evidence we gather and accept about student learning and how our interpretation of that evidence influences what we do in a classroom. Assessment often leads to innovation in the classroom, but innovations can create disequilibrium between student expectations and what occurs in a classroom. Many students still expect to be passive learners in an instructor-centered classroom where faculty talk and students listen (Ebert-May et al., 1997). Furthermore, the majority of beginning undergraduate students view knowledge as existing absolutely and concretely and assume knowledge to be certain (King & Kitchner, 1994). Students entering college are dualistic thinkers (Perry, 1970); they are intolerant of ambiguity; answers are right or wrong, black or white; they either get an idea or do not get an idea. Students are not prepared to try to understand complex or abstract ideas (Belenky, Clinchy, Goldberger, & Tarule, 1986). Therefore, it is not uncommon that faculty who are innovative and who attempt to create active learning in the classroom to

achieve higher-level thinking by students show a measurable decline in teaching evaluations, especially if the type of evaluation used has little to do with the reformed goals of the course. For example, student evaluations required for each class in the College of Arts and Sciences at my university include the following questions (Table 2).

The mean point values for these questions (based on a scale of 1-5) are tallied and individual faculty scores are compared to the college mean, usually 4.0 with a standard deviation of 1.0. Although faculty are encouraged to design and use other types of course evaluations, the reality is that the mean point score from this instrument becomes an important variable used during discussions about a person's teaching by promotion and tenure committees. Furthermore, cases have been reported in which a faculty member's mean score was 3.9 compared to the College mean of 4.0, therefore, this individual's teaching was considered below average. Faculty ignored interpretation of the mean and standard deviation in this case.

How does the student assessment instrument in Table 2 provide substantive feedback to faculty about the goodness of fit between student learning goals, pedagogical methods, and student learning styles? It does not. Furthermore, noninstructive negative feedback to faculty fails to encourage or reward thoughtful innovation and careful analysis about teaching and learning. It takes less time and involves fewer risks to teach in a traditional manner (with enthusiasm) and get higher point scores. We all know how to do that. Freire describes traditional education as "banking"—the instructor's role is "to 'fill' the students by making deposits of information which the teacher considers to constitute true knowledge" (Freire 1971, p.63). The students' job is merely to "store the deposits." He argued that traditional education does not enable learning, because students are never given an opportunity to liberate their mind and define their own essence.

An Alternative Solution

Noninstructive student course evaluations should be reconsidered. Students need to move towards a mode of reflecting and evaluating their own understandings and abilities and providing that feedback to faculty. We began utilizing student self-evaluations about course goals as a measure of learning and as substantive feedback to guide changes in faculty practice.

Self-evaluation is integrative, reflective work throughout a course that emerges as an ongoing process through various assessment strategies (Angelo and Cross 1993). Through this process, students reflect on and evaluate their own scientific understanding and ability. When students reflect on their accomplishments in a course, they are really conducting a self-evaluation. It is both a process—students think about, write about what they did and learned in a course—and a

Table 2. Student Evaluation: Course and Instructor Characteristics. For each item, students bubble in their choice on a scale of 1 (strongly disagree) to 5 (strongly agree):

1.	The objectives of the course were made clear to me.
2.	The instructor accomplished course objectives.
3.	The instructor seemed genuinely concerned with student progress.
4.	My interest in the subject has been stimulated by this instructor.
5.	The course was intellectually challenging.
6.	My general estimate of this course.
7.	General estimate of this instructor.

product—it is a written document that informs faculty about how students regard their accomplishments and how they achieved those accomplishments.

Student self-evaluation is both an old approach and a new one (MacGregor, 1993). The method was used historically in alternative colleges, and eventually more traditional colleges and universities began to use self-evaluation to engender students' active participation in the process of evaluating their learning. Learning theorists strongly advocate the value of having students think more reflectively about what they know and what they can do. As students gain experience, self evaluation becomes an important learning strategy as well as an avenue for alternative forms of assessment. While enriching learning for students, student self-evaluations also can help faculty and departments learn about student learning.

Feedback from student self-evaluations informs us about what keeps students motivated, engaged, and interested and about what they consider important in the course (MacGregor, 1993). What we read may provide us insight into the teaching and learning going on in our classrooms. Perhaps what we thought students learned, they did not learn at all; what we thought clear and simple was complex and confusing to students; what we predicted to be a successful pedagogical strategy was perceived as cumbersome by students (Kusnic & Finley, 1993).

Our research indicates that one of the most important effects of self-evaluation is providing a strategy that actively engages students in their learning, not only in terms of what they can do, but also in terms of what they cannot do at this point in time, what directions their learning must take, what must they do better. "Students who internalize valued achievement targets so thoroughly as to be able to confidently and completely evaluate their own and each other's work, almost automatically become better performers in their own right" (McMillan & Forsyth, 1991).

Self-evaluation: The Process

We used self-evaluation in a large introductory biology course for nonmajors with over 600 students enrolled. The self-evaluation was designed specifically for the students to comment on their accomplishment of the goals of the course. It was a written assessment in which the students provided both quantitative and qualitative synthesis of their learning. Students put themselves in the center of the learning experience, rather than focusing on the instructor. Writing self-evaluations was a challenge for students and required guidance, practice, and time.

We included the assignment and rationale for the self-evaluation in our syllabus. Students were informed on the first day of class that they would gain a perspective about where they had been as learners and knowers throughout the course and would need to consider what they should do next. To help students, we integrated formative reflective work throughout the course. For example, periodically we asked students to write short statements about the function of their cooperative groups. Alternatively, we would ask them to reflect on the effectiveness of an assessment strategy, such as concept maps, on their understanding. Importantly, we provided some type of feedback to student responses.

As a summative assessment, we asked students to provide us a self-evaluation at the end of the course that described their perception of their accomplishments in the course, as well as the accomplishments of the faculty. The instrument was based on student and faculty goals for the course, and each question had two parts. First, for each outcome students were asked to indicate on a five-point Likert scale the degree to which they accomplished each goal. Then in the space below each outcome, students were required to explain what happened during

the course that influenced the choice they circled.

The instrument was on the Web, and students had ten days to complete the assignment. Students electronically submitted the assignment, which included their names, and they were assured that we would not read the comments until final grades were assigned. Students received points for completing the self-evaluation, either all of the points or none. Maximum points were awarded to each student who completed the instrument fully, that is, written comments that included statements, examples and backing. No points were awarded if any information was missing or if a student merely repeated the question. A graduate student quickly examined each paper for completeness and awarded full credit or no credit. No students voiced concerns about anonymity since an environment of trust had been developed during the course. Furthermore, since the nature of the guiding questions for self-evaluations focused on the learner, students tended not to "blame" successes or failures on the instructor. The majority of comments focused on students' perceived accomplishments that often included likes or dislikes about the course in the context of their achievement.

Frequency of each response was tallied for each item. Then the written comments associated with each item were coded and interpreted using NUDIST software (Non Numerical Constructed Data, Indexing Search and Theory Building, QSR: Qualitative Solutions and Research, Inc., 1997). The combination of quantitative and qualitative responses provided two complementary approaches to interpreting trends in the data, a technique commonly used in science education research (Lancy, 1993).

Self-evaluation: Results and Interpretation

Figures 2-5 show the frequency of responses students provided regarding the degree to which they thought they achieved goals 2, 3, 4, and 8 (Table 3) for the course.

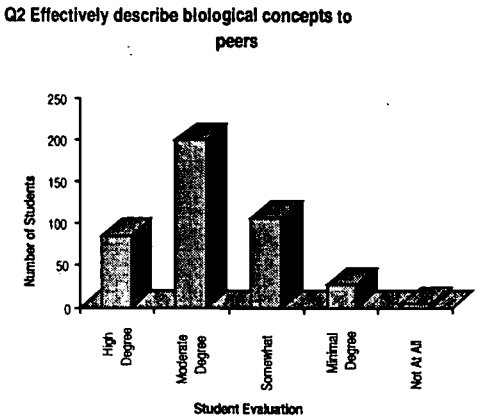
Table 3. Self-evaluation instrument for students in introductory biology. Each question includes a five-point Likert scale and an explanation section for extended responses as shown in Question 1.

1. To what degree have you increased your ability to describe how other people have used the process of science?
High Degree Moderate Degree Somewhat Minimal Degree Not at all
Explanation:
2. To what degree can you effectively communicate an understanding of and links among biological principles and concepts to peers and others?
3. To what degree have you developed confidence in your ability to write about, criticize and analyze concepts in biology?
4. To what degree have you increased your ability to use the process of scientific inquiry to think creatively and formulate questions about real-world problems?
5. To what degree have you developed positive attitudes about the relevance of biology to your life and the ability to apply this knowledge in the resolution of real-world problems?
6. To what degree have you enhanced your understanding of biological concepts and application of them to personal, public, and ethical issues?
7. To what degree have you enhanced your ability to reason logically and critically to evaluate information (i.e., be skeptical)?
8. To what degree did you develop positive interdependence and individual accountability within your cooperative groups?
9. To what degree did the instructor provide a learning environment in which all students participated in a variety of instructional strategies and assessment practices that challenged your higher order thinking and reasoning skills so you could successfully demonstrate the outcomes described above.

The most frequent student response was that they accomplished the goal to a "moderate degree," with over 90% of the students reporting they achieved the goal to a high degree, moderate degree or somewhat. We did not expect the majority of students to respond in the "high degree" category because the course was challenging to most students. Rather, we predicted that most students would recognize that they needed to know and do more to accomplish each goal to a high degree. The explanations students wrote after each choice confirmed this prediction. Students recognized they were making progress toward accomplishing the goal, yet, at the same time, realized that they needed more experience, practice, and knowledge to become highly effective in achieving the goal. Students were realistic about what they could accomplish in a one-semester course, but also described what they needed to continue to learn in the future.

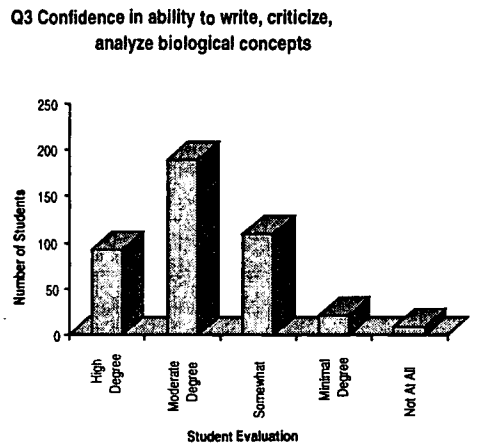
The sample quotations adjacent to Figures 2–5 represent the types of self-reflections commonly provided by students. We used these statements to help interpret the frequency distributions for each question, a combination of quantitative and qualitative assessment. For example, although all cooperative learning groups were not perfect (Figures 2 and 5), the majority of students wrote about the value of group work to their learning and provided specific examples of how their group functioned to achieve their goals. Alternatively, reasons for less effective cooperative groups were provided with specific examples. The comment in Figure 3 suggests that the extensive writing done in class was useful to students, although not necessarily their favorite task, "We were forced to write down how. . . ." Various performance assessments showed that students' writing improved significantly throughout the course. For example, students explained why the position papers and associated public hearings were a meaningful assessment strategy (Figure 4); therefore, we continued to use position papers and public hearings as an assessment strategy in the course.

Figure 2. Frequency of responses and sample explanation to the goal: To what degree can you effectively communicate an understanding of and links among biological principles and concepts to peers and others?



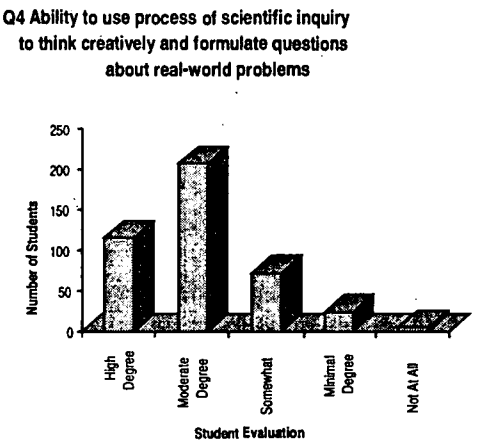
"So much group work made me realize that understanding a concept and being able to communicate a concept are different things."

igure 3. Frequency of responses and sample explanation to the goal: To what degree have you developed nfidence in your ability to write about, criticize and analyze concepts in biology?



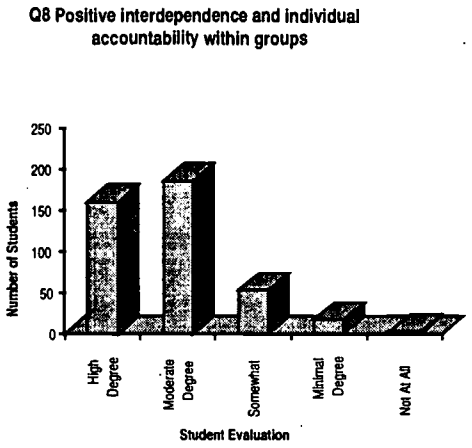
"I always was good at biology (or at least got good grades in biology), but never really understood any *connections*. This class made it clear to me that everything is connected to each other. We were forced to write down how we *understood* concepts, not simply to memorize parts and functions."

Figure 4. Frequency of responses and sample explanation to the goal: To what degree have you increased your ability to use the process of scientific inquiry to think creatively and formulate questions about real-world problems?



"We did our position papers on two real world issues that we deal with every day. One paper we did was on the environment and the other was on breast cancer. . . . This class really makes you think about how important scientists are in the world today. They try to formulate and answer questions that will help us survive in the future."

Figure 5. Frequency of responses to the goal: To what degree did you develop positive interdependence and individual accountability within your cooperative groups?



"Working in groups requires so much responsibility. This means coming to class every day, having your input for all of the questions and quizzes and most of all showing up on your own time to complete homework assignments, etc. . . . Working as a team requires dedication and cooperation This is how everyone will succeed."

In our course, we used student self-evaluation in the same way we required students to reflect about concepts when they compared and contrasted ideas, analyzed their work, or explored the implications of a theory. Both required reflective thinking. During the process of self-evaluation, students' learning moved from a passive process to an active, meaningful process (Kusnic & Finley, 1993). King and Kitchner (1993) consider evaluation one of the higher-level thinking skills that, along with analysis and synthesis, comprise "critical thinking." Hence, we incorporated self-evaluation into our course as one way to begin to move students along the continuum of intellectual development from concrete thinking to higher-level thinking. We considered self-evaluation a learning strategy that helped students construct meaning of concepts, derive relevance of ideas, and begin to build a coherent framework for continued learning.

Assessment is a learning process. Both faculty and students benefit from meaningful assessment information about the achievement of the broader course goals. Multiple assessment strategies can be implemented to provide evidence that students have or have not learned, have or have not accomplished the goals of the course. This feedback provides an instructor data to interpret and make informed decisions about student learning and teaching practice, similar to the process of data evaluation repeated daily in research laboratories. Self-evaluation is one strategy we used to reduce the risk of testing and refining innovative teaching practices, because it provided us direct information from the students about their perceived achievement of goals. Importantly, we used student self-evaluation as an alternative to poorly designed student evaluations that neither informed nor modeled best teaching practice.

References

Angelo, T., & Cross K. P. (1993). *Classroom assessment techniques: A handbook for college teachers.* San Francisco: Jossey-Bass.

Belenky, M. F., & Clinchy, B. M., Goldberger, N.R., & Tarule, J. M. (1986). *Women's ways of knowing.* New York: Harper Collins.

Brewer, C, & Ebert-May, D. (in press). Dynamic discussion in a large lecture hall: The case of genetic engineering. *Journal of College Science Teaching.*

Ebert-May, D., Brewer, C., & Allred, S. (1997). Innovation in large lectures—Teaching for active learning. *Bioscience, 47,* 601-607.

Freire, P. (1971). *Pedagogy of the oppressed.* New York: Seaview Press.

Hodson, D. (1992). Assessment of practical work. *Science and Education, 1,* 115-144.

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults.* San Francisco: Jossey-Bass.

Kusnic, E., & Finley, M. L. (1993). Student self-evaluation: An introduction and rationale. In J. MacGregor (Ed.), *Student self-evaluation: Fostering reflective learning.* San Francisco: Jossey-Bass.

Lancy, D. (1993). *Qualitative research in education: An introduction to the major traditions.* White Plains, NY: Longman.

MacGregor, J. (1993). *Student self-evaluation: fostering reflective learning.* San Francisco: Jossey-Bass.

McMillan, J., & Forsyth, D. R. (1991). What theories of motivation say about why learners learn. In R. J. Menges & D. Svinicki (Eds.), *College teaching: from theory to practice.* San Francisco: Jossey-Bass.

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

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn.* New York: Cambridge University Press.

Perry, W. G., Jr. (1970). *Forms of intellectual and ethical development in the college years: A scheme.* Troy, MO: Holt, Rinehart & Winston.

Moving the Mountain: Impediments to Change

Eric Mazur
Gordon McKay Professor of Applied Physics, Division of Applied Sciences
Professor, Department of Physics
Harvard University

There is no doubt that since the beginning of this century the United States has ranked first in generating outstanding scientists. It is therefore ironic that, as a whole, the population of the United States scores low in science and mathematics. One only need turn to the media to see that our society does not value science and science education as it did just a few decades ago. In spite of all the advances in science and the many contributions of related technological developments to society, science illiteracy is rampant. The average person has little faith in scientists, and more pressing problems than science education are on the agenda of most people. These developments are worrisome because for everyone to understand at least what science is about is in the interest of society. No one can deny the formidable advances that have been achieved in science and their impact on the quality of life—advances that would not have been made without the outstanding quality of American scientists. What happens now in the classrooms across the United States will directly affect the health and well-being of this country in the next century. We must act now to prevent losing our edge in science and technology.

At the college level, the introductory science course often is one of the biggest hurdles in the academic career of a student. For a sizable number of students the course leaves a permanent sense of frustration (Tobias, 1990). I only have to tell people I am a physicist to hear grumbling about high school or college physics—almost to the point of making me feel embarrassed about being a physicist. This general sense of frustration with introductory science is widespread among nonscience majors required to take science courses. Even science majors are frequently dissatisfied with their introductory courses, and a large fraction of students initially interested in science end up majoring in a different field. What have we done to make it that way, and can we do something about it?

Science education has been focused much too long on competitively generating a steady supply of future scientists. We must direct our science education not just at students going on to a scientific career but also at those majoring in other fields. It is time to realize that the demand for scientists is determined to a large extent by people for whom the introductory science course is the only direct exposure to science and who remember science only by the frustration it has caused them. It is time to realize that those who become successful scientists do so in spite of the current educational system, not because of it. It is time to realize that better science education for all will ultimately lead to a higher quality of life.

Broadening and improving science education will require a major change in attitude. The current mode of instruction is self-perpetuating: postsecondary faculty educate both their own successors and future secondary teachers; secondary teachers, in turn, prepare the next generation for a new cycle. At all levels one can find excellent teachers, but for the most part instruction in science is geared toward the scientist, not the general public.

The first step in remedying this problem is to create awareness: few faculty have a good understanding for what their students are actually learning in, let alone retaining from, their science courses. I, for one, had gone on lecturing happily for many years before

realizing that students were not at all learning what I wanted them to learn (McDermott, 1993). Students were memorizing by rote and learning to *cope*. My goal was to teach them physics; their goal was to get a good grade. How can one reconcile these two? An obvious answer is to make sure that the assessment—examinations, homework assignments, etc., all that enter into the final grade—properly reflects the goals of the course. Herein lies precisely the problem: the standard assessment is often a false indicator. Students often manage to score perfectly on standard problems without understanding any of the underlying basics (Mazur, 1997; McDermott, 1993). They have learned to solve problems mechanically, by memory or by analogy. A recent retention study carried out at Carnegie-Mellon has shown that, two years after completion of a traditionally taught introductory course, students' knowledge of the material is back to where it was *before* they took the course—the only things that remain of the course, of the students' and the instructor's efforts, are the final grade and memories that, in all likelihood, are not among the most pleasant ones. Clearly, before we can even begin to remedy any problems in education we must create a broader awareness of these problems.

One way to create broader awareness is to create assessment instruments that uncover failures in our current educational system. In my own field there currently is an abundance of such instruments, all of which focus on assessing students' understanding of important basic concepts. An excellent example is the *Force Concept Inventory* by I. Halloun and D. Hestenes. (Mazur, 1997, includes the most recent version of this test and other assessment instruments.) Still, by themselves, these instruments are not sufficient. It is all too easy for a skeptical instructor (of whom there are many) to dismiss the instrument as faulty—in other words to blame the test, and perhaps the students, instead of the method of instruction for any poor results.

What convinced me—and I counted myself among the skeptics—is so-called *paired-problem* testing. Instead of focusing on just the underlying concepts, for every topic taught in my course I began to ask a combination of two types of questions: one conceptual, the other traditional. The juxtaposition of these two types of problems is illuminating: what struck me when I began this paired-problem testing a few years ago was the enormous contrast in performance. My students did very well on the traditional problem, but poorly on what I considered to be a very simple qualitative problem *dealing with exactly the same basic concepts*. For the first time it became clear to me that students often simply recognized a traditional problem as one that pertained to a certain formula, in which case all they needed to do was to put the right numbers in the right place and work through the algebra—about the only thing necessary to solve the problem was the classification of the problem, the recognition of the correct equation or procedure. At the same time, their poor performance on the "simple" qualitative question was a clear indicator that the students had no clue what the equations or procedures they were using really meant. And, naturally, a few months after completion of the course, recollection of the equations and procedures fades away, leaving little substance behind. It is not sufficient to devise new assessment methods. A side-by-side comparison of students' performance on qualitative questions and on more traditional quantitative questions is a crucial element in uncovering the shortcomings of the traditional method of instruction and convincing faculty of the need to change.

Once awareness is created, the next step is to devise an instructional method that effectively addresses the shortcomings of the traditional method. I will not dwell on this topic, however, as my copanelists have extensively addressed this problem, and I have recently written on this subject (Mazur, 1997). Let me simply add that, with support from the National Science Foundation, I have

begun setting up a Web site aimed specifically at disseminating successful and simple-to-implement instructional practices *http://galileo.harvard.edu.*

A final point I wish to address in this paper is what I believe to be an important impediment to change. Failure to recognize this barrier seriously compromises our efforts. Put simply, the problem is the following: the initial effect of any change is not an improvement, but a period of problems, adjustments, mismatch, and to some extent a period of frustration and pain. A good analogy is that of a tennis coach who discovers that one of her students is not holding his racket correctly. After adopting the correct grip, the student's performance does not immediately improve—in fact, it gets *worse*: most balls go into the net or off to another court. The student gets frustrated because he is used to holding his racket differently and, it now appears, played *better* the old way. The coach, however, knows that the student can never improve without the correct grip and that her student's play will soon improve.*

An instructor of a large introductory class is likely to think differently when facing the often intimidating discontent of his audience. "Am I doing something *wrong*?" or "This is not working!" are natural reactions. After I changed my method of instruction, one student asked me, "Professor Mazur, when are we going to do some *real* physics?" Others, having done well in their high-school physics class, are very disgruntled when they discover that their high-school performance does not translate into a good performance in my class—they blame their performance on this so-called improved method of instruction forced upon them by me. Expecting immediate improvement is not only naive, but is bound to lead to disappointment. It is therefore important to make instructors who are embarking on a new road aware of this obstacle. In addition, it is important to prepare the students for this transition.

An additional problem is that the keeping of statistics and the desire to maintain a "fair" and consistent grading system from one year to the next tend to perpetuate the status quo. An administrator once told me, "Your method does not work because our good students no longer do well—look, there is very poor correlation between your results and our statistical analysis of course performance over the past ten years!" "Good" here means "good by the traditional standard," and this is the essence of the problem: changing the method of instruction also means changing the method of assessment. How else can one assess the success of the change and, what is perhaps even more important, drive students to change? Changing the method of assessment, however, means giving up any meaningful correlation with previous assessments. As long as administrators and faculty do not realize that this poor correlation is an unavoidable consequence of change, it will be impossible to move forward.

References

Mazur, E. (1997). *Peer instruction: A user's manual.* Upper Saddle River, NJ: Prentice Hall.

McDermott, L. (1993). How we teach and how students learn—A mismatch? *American Journal of Physics, 60*, 295).

Tobias, S. (1990). *They're not dumb, they're different: Stalking the second tier.* Tucson, AZ: Research Corporation.

* This wonderful analogy was pointed out to me by Phil Sadler of Harvard University, who recently studied student performance and grade correlations in classes using novel approaches to teaching.

The Integrality of Assessment

David B. Porter[*]
U.S. Air Force Academy

Before addressing the questions posed for this session, I want you to understand a little about the context and perspective in which my ideas developed. Although each institution of higher education is unique, federal military academies are, in some respects, in a category of their own. Many of the fundamentals of teaching, learning, and motivation, however, are likely to be similar to those encountered at other colleges and universities. The U.S. Air Force Academy is a four-year undergraduate university. Its mission is to develop and inspire air and space leaders with vision for tomorrow. Academy cadets (students) are competitively selected and do not pay tuition. They incur a commitment to serve as commissioned officers in the Air Force for five years after graduation. The size of the Cadet Wing (student body) is about 4,000; about 14% are women and 18% represent racial minorities. As a group, cadets are bright (average SAT scores are typically above 1300), athletic (more than 80% earned letters in varsity sports in high school), and tend to share conservative social and political perspectives.

The academic faculty of 530 is about 20% civilians and 80% military officers assigned to 19 academic departments in four academic divisions: Basic Sciences, Engineering, Social Sciences, and Humanities. Since an assignment to the Air Force Academy is considered a special duty for Air Force officers, most military faculty members serve only a single three-year tour. As a result, the average teaching experience among faculty is only 2 years, and annual turnover in the academic departments often is 30%. Recent hiring of full-time civilian faculty has increased the proportion of faculty with doctorate degrees to about 50%. Class sizes are usually less than 20 students, and an average teaching load for junior faculty is 4 sections (12 semester hours). Teaching and learning are emphasized across the faculty. Several years ago, governmental emphasis on quality increased the focus on assessment as a way to enhance institutional effectiveness.

Student Learning: How can faculty best understand what undergraduates are learning, retaining, and using in future contexts?

Few engineers would entertain the notion of building a bridge contrary to the laws of physics. Although somewhat less precise, the relationships that gird the mental world of thinking and learning are just as potent as those that constrain the physical world. Faculty who design and deliver curricula without understanding the principles of human learning are likely to waste their own time and create classes that harm students.

The process of learning is susceptible to the same scientific method used in the physical sciences. A coherent model of human learning is a necessary starting point for such inquiry. Learning has adaptive significance; it is a natural phenomenon for humans; it enhances our individual and collective chances of survival (Gould, 1981). Learning assumes that certain experiences and activities enhance the capacity of the individual to deal with environmental challenges. Response quality (i.e., performance) can be influenced by many things (viz., knowledge, skills, and attitudes).

[*] Opinions presented in this paper are those of the author and do not necessarily reflect those of the USAF Academy, Department of Defense, or any other government agency.

Changes in any of these components can affect both performance and learning. Thus the question of student learning might be more accurately considered as several questions:

What is known that wasn't known before?
What can be done that couldn't be done before? and
What is the effect on student attitudes? (Porter, 1991).

The relationships among these components of learning are even more important than the components themselves. For example, the type of mental activity involved in study (i.e., a skill) is a much better predictor of retention than is duration of exposure (Craik & Lockhart, 1973; Craik & Tulving, 1975). Elaborative rehearsal typically involves activities such as reflection, comparison, argument and conclusion. Maintenance rehearsal is the act of simply repeating a phrase, formula, or particular "fact" verbatim. Elaborative rehearsal usually yields two to three times greater retention than maintenance rehearsal. Ironically, the most common student study strategy is to "bear down" and rely heavily on maintenance rehearsal, especially in those subjects students find most challenging or distasteful (i.e., science and mathematics).

Another effective but underemployed strategy is visualization. Although there are individual differences, material that has been visualized is about twice as likely to be recalled as material that has been verbalized for an equal amount of time (Atkinson & Raugh, 1975). In fact, there is strong evidence that the greatest retention is likely to occur when presentations and activities involve both visual and verbal processing (Pavio, 1971). One final example of how process affects knowledge retention is known as the self-referent effect (Rogers, Kuiper, & Kirker, 1977). Students remember best what they care about most and what connects to them most personally. In a typical experiment, subjects are asked to rate one list of adjectives on a 5-point scale ranging from positive to negative. Students rate another list of adjectives on a 5-point scale ranging from "most like me" to "least like me." The usual result is that subjects recall three times as many adjectives that they rated in relationship to themselves.

Elaboration, visualization, and self-reference are not innate study techniques. In fact, many students (and faculty) steeped in an academic version of the Protestant work ethic (viz., no pain, no gain) assume such techniques are simply frivolous diversions. To enhance learning, these techniques must be presented, advocated, practiced, and actively incorporated into lessons and courses with cooperation and trust between students and their teacher. To the extent students are fearful of failing, see the teacher as an adversary, or are engaged in direct interpersonal competition, the necessary collegial classroom climate is unlikely to develop (Kohn, 1986; Glasser, 1990; Palmer, 1998). Recognizing the importance of student attitudes and driving out fear are often prerequisites of pedagogical progress and development. Educational success is often contingent on students seeing faculty as allies in their battle against a common enemy, ignorance.

However, all these techniques still do not directly address the question of "how" faculty can best understand student learning. As Parker Palmer suggests (1998), technique is what one uses until a real teacher shows up. While mastery of techniques may be necessary, no level of mastery is sufficient to assure insight. After several years of working with good people, who strongly desire to become effective educators, I'm convinced the single most critical variable is also one of the least tangible, authenticity. Teachers must be able to be themselves; "who they are" and "what really matters to them" are what create the classroom climate and provide the crucible in which learning might occur. Students need a safe place to hold substantive conversations before they will share the secrets of their private perceptions and assumptions. Within a supportive context, I've found a single three-word phrase

to be very helpful in drawing out the information I need to understand what students are learning. These three words are, "Help me understand." However, as a mere technique even these "magic words" are likely to be ineffective if the teacher doesn't really mean them or hasn't yet convinced the students of this. The words must "fit" the teacher and also the rest of the course; they must be "authentic."

Students report the greatest learning when faculty emphasize all three types of outcomes (knowledge, skills, and attitudes). A study of 115 Air Force Academy faculty members suggested teachers who balance emphasis on students' knowledge, skills, and attitudes increase students' subsequent perceptions of their own learning. In fact, this effect was larger than the influence of teaching experience, teacher temperament, and degree-level combined (Porter & Benson, 1995).

As part of the Air Force Academy's recent effort to assess the contribution of 35 core courses to these three educational outcomes, faculty teaching core courses were asked to rate the emphasis placed on three kinds of educational outcomes: knowledge, critical thinking skills, and intellectual curiosity. Emphasis on integrated fundamental knowledge was relatively equal across all four academic divisions. Unit-weighting the emphasis on knowledge made divisional differences in relative emphasis on critical thinking and intellectual curiosity more apparent (Porter, 1997). Table 1 shows, on the left, that faculty teaching Basic Science and Engineering core courses reported placing much less relative emphasis on student skills and attitudes than did faculty teaching Social Sciences and Humanities. The ratings shown on the right of the table were provided by faculty assessment teams as part of a comprehensive assessment effort. Seven interdisciplinary faculty teams considered a wide range of standardized inputs from students, faculty, and course syllabi to determine each course's contributions to students' mastery of integrated knowledge, ability to frame and resolve ill-defined problems, and intellectual curiosity (Porter, 1997). Contributions were rated on an absolute 7-point scale with 4 being neutral.

Since courses were assessed separately, the results could also be used to identify which pedagogical practices were associated with contributions to the three outcomes. The extent to which a course involved group work was negatively associated with the course's contribution to students' attainment of integrated knowledge and increase in intellectual curiosity. Closer examination showed that this was especially true in engineering core courses where the correlation between the proportion of group work and assessed contribution to students' knowledge approached -.80. The proportion of the student's course grade that depended on computation was negatively related to its critical thinking contribution (but showed a slightly positive relationship to knowledge attainment). Technical courses specially designed for nontechnical cadets were generally found to contribute the least of all core courses to any of the three outcomes; these were also the courses that tended to employ student groups the most (Porter, 1997).

Table 1

	Relative Emphasis (based on faculty self-reports)			Average Rated Contribution to: (7-pt scale - 7 assessment teams)		
	Knowledge	Skills	Attitudes	Knowledge	Skills	Attitudes
Basic Sciences	1.0	54%	63%	4.88	4.66	4.68
Engineering	1.0	85%	69%	5.03	4.40	4.36
Social Sciences	1.0	94%	87%	5.36	5.56	5.42
Humanities	1.0	114%	108%	4.79	5.06	4.94

Other pedagogical process variables have also been found to relate to student learning. Although competition broadens performance distributions and allows selection among students to occur with greater confidence, its net effect on learning is generally negative (Kohn, 1986). Students also learn best when they are allowed to make mistakes, identify, and correct them. Overemphasis on external contingencies (either rewards or punishments) is likely to leech satisfaction and pride of ownership from learning (Glasser, 1990). Ironically, learning becomes much more likely when faculty simply "lighten up" and "let it happen." In classrooms with coercive climates, data about student learning are likely to be resisted and resented. In contrast, in classes with more collegial climates, feedback is likely to be used to enhance understanding, on both sides of the podium.

Class Content: How can faculty best judge the utility of their class content choices? (How do we know if we are teaching the right stuff?)

In my opinion, "how to teach" is a much more significant question than "what to teach." In fact, once the how is mastered, practically anything can be taught and learned. However, what to teach does matter, and careful consideration from many perspectives is appropriate. Material should be relevant—the more salient the connections to "real life," the more likely the material will be intrinsically motivating to students (and the less coercion will be required to get them to study). Course content should connect to what students already know as well as what they desire to learn. To the extent students perceive the course material as a bridge between their present situation and the attainment of their aspirations, little external pressure is required to motivate them to study. This is not a recommendation to "lower standards," "reduce rigor," or "pander to the lowest common denominator," Students take pride in accomplishing challenging tasks if they recognize the relevance of the material and know they'll receive the support needed to be successful. The particular content of courses should be determined by the current understanding within the discipline and its range of applications to business, industry, and education. In fact, efforts to do this at the Air Force Academy have paid big dividends in terms of student motivation. Faculty sharing personal experiences showing the relation between the topics covered in class and Air Force operations often heightens students' intrinsic motivation considerably.

In order to appreciate the potential contributions of particular disciplines and perspectives, students need to learn the "stories" that structure the discipline. For this reason, teachers who are naturally inclined to consider "the big picture" rather than prematurely focus on details are likely to be seen as being more effective educators by their students (Porter & Benson, 1995). This preference is reflected by the iNtuitive vs. Sensing dimension of the Myers Briggs Type Indicator. The correlation between a faculty member's preference for intuition over sensation and rated effectiveness is about .30. In comparison, differences in teachers' Introversion or Extroversion preferences show nearly no association with students' assessment of teacher effectiveness or students' learning.

Pedagogical Issues: How can faculty assess the goodness of fit between student learning goals, pedagogical methods, and student learning styles? (What kinds of classroom feedback best guide changes in faculty teaching?)

It is essential that institutional purpose, policy, and pedagogy align with one another (Porter & Light, 1994). For example, if the purpose is *development*, then collaborative approaches that encourage students to teach one another are appropriate. However, if the purpose is actually *selection* through

differential performance, then competition is a more appropriate paradigm. Faculty themselves must have a coherent story; they should know what they want to accomplish and should have developed a range of activities that support the accomplishment of these objectives (Angelo & Cross, 1993). They must also regularly check progress toward these goals.

Education is not a ballistic process; it is much more like sequential hypothesis testing. Each lesson plan is a hypothesis; it is derived from the teacher's mental model of the way the world (or at least the classroom) works. During the execution of the plan, feedback from students should be collected and considered frequently. Classroom assessment techniques, such as the three main (or "murkiest") points, quizzes, verbal responses, and student questions, all provide relevant information (Angelo & Cross, 1993). Most of us need to adjust our models as evidence disconfirms some of our initial assumptions. If we truly endeavor to model learning, making adjustments should be a cause for celebration rather than chagrin.

It is important to recognize and appreciate that individuals differ from one another. There are likely to be at least several viable approaches to any particular problem, especially ones relevant to the "real world." A diversity of perspectives can enliven classroom discussions and enrich learning. Diversity creates a constant tension and provides the impetus for students and faculty to engage in collegial conversations with the goal of understanding each other's perspective.

However, the influence of individual learning style differences on most general educational outcomes is very small. It might even be argued that it is more important that students develop skills contrary to their natural affinities. Fortunately, faculty do not have to develop individual syllabi for each student. What they can do is create a syllabus rich with options and opportunities for students to master the material and acquire important individual and group skills in the process. The positive effects of providing choices and respecting a variety of student abilities are far greater than precisely matching academic tasks to student temperaments. Late adolescence is a volatile time; the reliability of most learning style instruments does not support reliance on categorizations determined early in the semester, let alone those acquired during freshman orientation.

Risks and Costs of Innovation: What are the risks (apparent or real) of classroom innovation? What does it take for individuals to undertake them? How can assessment reduce them? What are the costs of classroom innovation? Who bears them? How can assessment address them?

Every system is perfectly designed to yield the results observed. If perceived risks are sufficient to stifle innovation, the institution's days are numbered. Competition among educational institutions is increasing; there are more routes to a wider variety of degrees than ever; institutions that do not adapt will wither. *To innovate or not to innovate,* is that really the question? If an innovation succeeds, teachers, students, and the institution win. If an innovation fails, valuable information is gained and once again students, faculty, and the institution benefit. However, if innovation is absent, nothing is learned; alternatives disappear and everyone loses. Innovation creates the variability necessary to assess effects, consider alternatives, understand processes, and improve institutions.

However, innovation for its own sake can be dangerous: there is a difference between a vision and a hallucination. Effective assessment can help distinguish the two (Porter, 1997). Bureaucratic traditions that developed in an environment of shrinking budgets and intramural competition encourage inappropriate and ineffective approaches to assessment. Certain cosmetic approaches to assessment attempt to insure positive appearances at the expense of

gaining insight into processes and effectiveness. Such pseudo-assessments hide variance and obscure causality in their rush to proclaim perfection. The sine qua non of assessment is the same as for science: we must endeavor to disprove our own assumptions. In an educational community, where trust and understanding are valued and individuals are respected for their commitment to the common enterprise of learning, institutional and classroom assessment can contribute to the educational process. Interestingly, the keys to creating learning organizations are very similar to those for enhancing student learning: authenticity, understanding, and trust.

References

Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers*. San Francisco: Jossey-Bass.

Atkinson, R. C., & Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory, 104*, 126-133.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology, General, 104*, 268-294.

Glasser, W. (1990). *The quality school: Managing students without coercion*. New York: Harper & Row.

Gould, S. J. (1981). *The mismeasure of man*. New York: Penguin Books.

Kohn, A. (1986, September). How to succeed without even vying. *Psychology Today, 20*, 22-28.

Kosslyn, S. M. (1981). The medium and the message in mental imagery. *Psychological Review, 88*, 46-66.

Palmer, P. (1998). *The courage to teach*. San Francisco: Jossey-Bass.

Pavio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rhinehart and Winston.

Porter, D. B. (1991). A perspective on college learning. *Journal of College Reading and Learning, 24*(1), 1-15.

Porter D. B. (1997). *United States Air Force Academy Educational Outcomes Assessment Working Group, Phase 2: Final Report* (Technical Report 97-2). Colorado Springs, CO: USAF Academy

Porter, D. B., & Benson, M. J. (1995). Correlates of course and faculty perceived effectiveness. In D. Porter (Ed.), *United States Air Force Academy Educational Outcomes Assessment Working Group, Phase 1: Final Report* (Technical Report 96-4). Colorado Springs, CO: USAF Academy.

Porter, D. B., & Light, J. (1994). The institutional integrity of educational systems. In *9th Annual Conference on Quality and Assessment in Higher Education, Commissioned Papers*. Washington, DC: American Association of Higher Education.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology, 35*, 677-688.

Panel 2:
Assessment and the Promotion of Change in Departments, Disciplines, and Institutions

The Reaction to the Symptoms Versus Reaction to the Disease

Brian P. Coppola
Department of Chemistry
The University of Michigan

Recognizing and Respecting Different Values

I first heard Nobel Laureate Arthur Kornberg tell this allegorical tale in 1989, and I have found it to be broadly useful over the years.

> A physician, jogging along the beach, is shocked to see large numbers of people running from the shore into the water . . . and then drowning. He rips off his jogging suit and dives right in, dragging the first victim back to shore. As he successfully resuscitates this individual, he is dismayed to watch as ten more people have run into the water. He is exhausted by the time he has dragged his fifth victim back to shore, and dismayed beyond belief to see the first person he saved stumble back towards the water. Off in the distance, he sees one of his colleagues from the medical school sitting on a piece of driftwood, just sitting . . . watching the situation unfold.
>
> The original physician yells out, "Say, you there! Doctor! What is the matter with you? Why aren't you helping me save these people?"
>
> "I am," she replies. "I'm trying to figure out why they're running into the water in the first place."

I like this story because it reminds me that there are many roles for educators to take, and all of them are indispensable. We need faculty who are our "general practitioners," front-line interventionists who, by virtue of reflective practice and intimate, long-term contact with students, anchor all of us in the realities of classroom instruction. The investigations conducted by these faculty are sometimes called "action research," that is, using immediate observations about student learning to make extemporaneous adjustments to improve instruction (perhaps it is just "good teaching"). We also need faculty who take stewardship roles for the system. We need people who can see in a more collective way the broad directions being taken by individuals in the context of historical, sociological, and philosophical trends. In this latter category, there is a significant responsibility for individuals not to be swayed by their own idiosyncratic situations (including "making a name" for themselves) and to keep these separate from whatever meta-analyses are conducted. Like the Kornberg story, progress relies on both groups doing their job responsibly. We do not always recognize the independent value of these roles, however. Faculty who do extraordinary jobs in one-on-one and small classroom settings are increasingly being required to adopt the practices of their more resource-rich colleagues in larger institutions. Faculty at larger institutions are also pressed to emulate styles of

activities that are sometimes better suited to the smaller schools with long-standing traditions in liberal arts education. For instance, group learning strategies that were developed in junior-level writing seminars for 15 students should not necessarily transfer into first-year science classes with 350 students without the faculty first reflecting deeply about how understanding the subject matter is benefited beyond just changing the affective dimensions ("I did not learn any more, but I liked going to class"). There are large institutions, both public and private, where the commitment to small class sizes has simply created an underclass teaching staff. In some places the strategy seems to work because most everyone is happy: the "research faculty" get to do their work, the "teaching staff" get to teach, and the students get caring instructors. But this solution also disintegrates American higher education at its core, and only a better education should drive such changes. In other arenas, faculty involved in large-scale projects are sometimes behaving more like managers who seek out subcontractors instead of being selected as creative visionaries who can identify and set broad conceptual directions. Sitting at the intersection of these scenarios is the emergent need for accountability that has accompanied all sorts of funding for curriculum development and reform over the last decade. This is not to say that I favor a return to an ascriptive culture of entitlement, but rather I argue that our current culture of accountability has not required much more than the appearance of substance. When assessment of student learning is done well, it should always provide for formative improvement to instruction by providing rich information. Not every assessment strategy provides this kind of information and, after all, not every experiment produces positive results. Creative people need to be wary of spending more time worrying how to affirm, prove, or justify that what they do is creative compared with evolving their teaching in creative ways. Meaningful collaborations between innovative faculty whose expertise is in the subject matter and those in the areas of education science work to the benefit of all parties. And that brings me to the topic of this NISE Forum: assessment. To a great degree, the demand for accountability has driven the current assessment movement.

Recognizing and Respecting Different Evidence

SMET faculty are accustomed to evidence that fits the usual criteria for physical science: we conduct procedures that can be replicated, and we create objective information, especially measurements, that can be reproduced. These measurements are assessments. Evaluation of these data, on the other hand, is strictly interpretive. In fact, progress in science relies on the debate between interpretations. The nature and context of assessment outside of naturalistic phenomena is distinctly different from melting points and tensile strengths. In educational reform, it is an error to treat assessment as naturalistic evidence. Furthermore, any given assessment does not leverage systemic change any more than any single data point can define a trend. Our focus on assessment currently serves many purposes: one is as a basis for argument and persuasion, another is classroom research, and another is a vehicle of accountability. All of these are peripheral distractions, or what I have called reactions to the symptoms of the challenges that face higher education.

Moving Past the Symptoms

In the twentieth century, the demands of an increasingly industrial and technological society dramatically changed the intrinsic nature of American higher education. The need for technical and professional training took precedence over the cultivation of virtue and the broader connections to what we now rather arrogantly define as the "nonscientific" parts of higher education. The German model for higher education supplanted the ecclesiastical one that had

been established in most of the United States, and colleges and universities became places where one could receive a specialized education leading to a career in such fields as engineering and science. We all recognize the rapid rate of advancement that has occurred in the technical and scientific disciplines. This century opened fresh on the heels of the Industrial Revolution and with the emergence of the engineering and physical science disciplines. In the United States, two World Wars, a space race imbedded in a Cold War, and the institution of federal public funding only further accelerated the rate of direct and indirect (spin-off) technological developments. The obligations for educating undergraduate and graduate students in science is wholly owned by the science faculty. Every precollege science teacher and all future faculty members are in our introductory science courses. Some of these individuals take advanced courses, and fewer still join our research groups. Science faculty are the sole caretakers for what constitutes acceptable practice in the educational and professional development of students in science courses, regardless of whether they are the minority who become scientists or the majority whose formative understanding and attitudes about science rest on these classroom experiences.

There are philosophical discrepancies (the "disease" from my title) at the core of science education today that need to be addressed, and good methods of assessment will certainly be crucial for addressing these problems. The vast majority of faculty earn their salaries by receiving a teaching assignment from their institution. On the other hand, even the appearance of excellence in undergraduate instruction can be automatically (and rather irrationally) attributed to inattention to one's research program. Do good teaching but do not do too well, and whatever happens do not get a teaching award. This is one of many conundrums facing young science faculty, in particular. The most successful, independent, and self-motivated graduate students from the most active research groups in the top-20 institutions become faculty, and the situation of their graduate department hardly ever matches that in which they find themselves. Some fraction of new faculty, I suspect, are selected precisely because they lacked any need for mentoring or education in the broader aspects of a life in science: they matched perfectly the prevailing culture in their graduate program. No wonder there is such a dramatic period of adjustment, something one of my colleagues calls the "assistant professor syndrome," where organizing and motivating the behavior of young scientists who are quite unlike the new research director becomes the task. The challenge of mastering these significant responsibilities comes as a surprise to new faculty members and takes place alongside the formidable task of developing an independent, international identity within a five or six-year time period. The scientific training of future faculty neglects most of these broader professional development issues. A new faculty member should not have to invest so much time simply learning how to do these things, because it automatically reduces the available time for actually getting the work done. On top of these demands, this new faculty member will also be assigned to organize and carry out instruction in undergraduate and graduate courses, the preparation for which is an area nearly neglected during graduate school. Graduate student teaching assistantships in science are remarkably different from those in the rest of the university. Unlike many disciplines, we use our students when they are least experienced; we do not invest them with decision-making responsibilities about what they are teaching; and the majority of programs provide little in the way of guidance beyond survival strategies for being in the classroom. Whether it is the first time these individuals are assigned to an introductory graduate course or to an undergraduate course, as new faculty members their most common teaching strategy is not at all surprising: "Who has a

good set of notes for this course?" If we do not provide as much training as we can for the demands of a professional life in higher education, then we put our young faculty into situations where they must make choices that would not be necessary with more appropriate preparation.

As a faculty, scientists work at cross purposes regarding the goals, the nature, and the implementation of a scientific education. Our idea of a syllabus is a list of subject matter topics and the order in which they are covered. When asked to generate a list of instructional goals, we produce the same document. In public forums, we speak to the need for an educated citizenry capable of critical analysis, but knowing and testing the items on the list is what we ask for. Our own technocratic education leaves us unable to make even simple connections to the rest of higher education (sociological, philosophical, literary, and so on) until after we are assigned to teach the "nonmajors" course. There are laboratory instructors who consider it simply wrong to put beginning students in a situation where the outcome is certain, and others who think that these instructors are wrong, although it still takes too long to realize this is the nature of the debate because there are two lines of dialog that can comprise discussions of pedagogy: Faculty member A says, "I'm right." Then faculty member B either says, "I'm right," or "You're wrong." Do not misunderstand my intent with these examples. I am not advocating that everyone needs to do the same thing the same way. I am recommending that the narrow band of activity called assessment cannot exist outside of the broader instructional culture. In order for assessments to have meaning, the information must be evaluated. In order to be evaluated, some notion of the relative value for different outcomes must exist. What are the relative values for (a) memorizing and returning long lists of information, (b) generating correct numerical solutions, (c) providing interpretive text, (d) analyzing new and unfamiliar information from the primary literature, (e) conducting a laboratory manipulation properly the first (second, third...) time in order to achieve a prescribed result, (f) repeating a self-generated laboratory procedure to demonstrate improvement, and so on? The answer might be that they are all equal in value. If this is so, then does the course place congruent emphasis on ensuring all outcomes? How does that happen? Suppose one is valued in the rhetoric and another is valued in the practice?

Categories for Assessing Student and Faculty Performance

The question of assessment extends from student learning within the course to faculty performance in its design and implementation. These are linked. In my own work, I have suggested that there are at least six categories for assessing student and faculty performance that differ in their intent and in fundamental information provided. In all of these categories, there are ranges of options, designs, and effectiveness that are possible, but my purpose here is to identify categories and not to look at any specific elaborations.

The first three categories constitute traditional classroom techniques organized by the faculty member or instructor in a course.

Category 1: Examinations, papers, reports, and projects.

These are still a primary mode of information used by instructors. The evaluation of these assessments is strictly product-based. Because of the design of these assessments, we assume that we can infer something about the process by which these materials are constructed. Alternatively, we may be satisfied that ability to replicate an acceptable performance by any route is acceptable, although it is incorrect to conclude that the appearance of any product implies a unique or even correct path. In

product-based assessment, even what appears to be a correct or acceptable presentation is primarily an interpretation of the instructor, because the artifacts that indicate the student's process are missing.

Category 2: Periodic sampling of intermediate materials.

There are many different versions of this. Many of our colleagues who seek to monitor the development of student learning use "drafts" or "studio practice" in order to customize interventions as needed. Writers of all kinds use multiple drafts to help sort out intermediate ideas (including science faculty, when you think about it, once writing manuscripts and grant proposals becomes necessary). Nearly all of the so-called active learning strategies are ways for instructors to get feedback on the intermediate learning of students. From the methods for "classroom research" popularized by Cross and Angelo, among others, to a variety of written and oral in-class Socratic instruction techniques, these assessments allow an instructor a more explicit look at the intermediate understandings of learners. These assessments treat the classroom setting as a kind of extended conversation, where the instructor is asking the questions Do you understand what I am saying? and Can you tell me how you know?, and the students are replying through various means. The evaluation of these assessments includes path as well as product. Portfolios are a common way to present assessment in this category.

Category 3: Peer-based editing.

An instructor who collects products or drafts of intermediate materials will often use private criteria to evaluate student work. Another objective can be to provide instruction in assessment and evaluation criteria along with the subject matter lessons so that students can begin to more meaningfully take on the instructor's role when working with their peers. If we want students to improve their critical skills for self-assessment, then they can learn a great deal from how differently their peers will approach the same assignment or task. Just as is true for proofreading your own work, at some point your "internal editors" fail and you rely on your "external editors" to help refine your understanding in addition to the way it is expressed. All of the assessments in the first and second categories actually represent teaching events, where a person is called upon to express understanding. This fundamental reversal in the student-teacher roles is not widely appreciated. In fact, you learn quite differently when you are explicitly aware that you are going to have to teach what you are learning. (Does that seem familiar to you? Every faculty member shares at least two common cultural heritages: we go around during the first few years saying "I never really learned this until I taught it" and we consistently encourage students to work together.) By creating opportunities for structured peer review and critique, where we teach the criteria for assessment and evaluation along with the subject matter, we can make assessment more public. The evaluation of this sort of assessment requires that students are guided in reflective practice. An instructor can use the record of a peer evaluator's work (reviewing the review) as a basis for understanding that student's abilities.

The final three categories are more formal research areas that are not typically part of the science education tradition. These often benefit from or even require a collaboration with someone in education science or cognitive science. In all three categories, longitudinal studies can often reveal the most important trends, the term- and year-long studies are much easier to manage and also provide useful information.

Category 4: Performance-based assessment.

Under the supervision of a facilitator, monitor, or recorder, student subjects participate in a task that is meant to evoke an aspect of learning from the course. Laboratory students might be faced with an unknown substance and they must think aloud about how they would learn what it is, all the while being tape- or video-recorded so that a series of researchers can code and analyze the strategic process used to solve the problem. Many expert-novice studies involve performance-based assessment, where the skills of a subject group, or perhaps different student groups, are evaluated by comparing student work with how faculty and graduate students solved the same problem.

Category 5: Large-scale survey work.

Statistical analysis of self-reported, scaled responses is complicated social research. Absolute differences observed in pre- and posttest designs have to be carefully controlled for pre-existing variations in the population(s). Collecting multiple responses related to a given attribute and gathering them together to form a variable is desirable. Interpreting numerical results relies on a number of factors: whether the observed changes are self-consistent within a certain theoretical framework as well as whether the changes make sense within the context of the course or intervention in question. It is not surprising that the course evaluation surveys typically collected at the end of the term are highly criticized.

Category 6: Interviews, observations, and focus groups.

Rich anthropological studies can reveal some of the most important information about student learning and faculty instruction. Such term-long and year-long studies of change rely on direct interaction with students and faculty by a third party who observes the process of instruction and learning as well as interviews the participants . These studies are intensive on all counts and are the rarest kind of information we have about science education. Small-scale efforts in this category include peer review by classroom visitation and one-time midterm focus groups conducted by an external evaluator (SGID: Small Group Instructional Diagnosis, for example). Evaluation in this category depends on what kinds of questions have driven the dimensions along which information is collected, which in turn are a result of a theoretical framework. For example, there are generalizations about the changes that might be expected within a student population or the way in which classroom instruction should be congruent with course goals, assignments, testing and grading.

Different things can be learned from each of these categories that can complement and inform information from the others. Assessment across multiple dimensions allows us to understand teaching and learning better than assessment from only one or two of them. But I return to the issue I raised earlier: good assessment is not enough. Assessment information is relatively neutral, but its evaluation is intractably linked to every other aspect of the instructional setting, from goals and implementation to infrastructure and rewards. The greater the incongruities within the instructional setting, the more difficult it becomes to make conclusions and recommendations.

University faculty, outside of schools of education, are notorious for their disdain of pedagogy. As scholars we seem to feel that knowledge of content is all that matters. If we provide a good course, full of the latest developments in our field, students will learn. We focus on teaching rather than learning, often with disastrous results. Lunch table conversations about how our courses are going are filled with destructive nostalgia about how much better students

were "in the old days." Facilitating a broad-scale conversation about pedagogy is a difficult task, particularly in a research university where faculty are engaged in exciting scholarship, but a morally reflective educational practice (which is a type of content) demands that pedagogy be taken as seriously as factual content. At least in the public eye, students are the reason for the existence of the university. Their interests in a high-quality education that prepares them to be effective participants in the society are paramount. We must move beyond the views that (1) teaching is merely the organization and delivery of content and (2) the primary goal of pedagogical innovation is the production of "artifacts" such as textbooks or, currently, interactive computer programs.

Pedagogical innovation requires changes in faculty behavior, the most difficult change of all. It is the difference between knowing (intellectually) that a good diet and regular program of exercise are truly the right things to do and the observation that the world has plenty of overweight, sedentary physicians who also smoke. Understanding the right thing to do is altogether different from having the will to carry it out, even when it is your responsibility to provide the education in the first place. Because this change must come at the core, the process will be slow. The first step is to facilitate a public discussion of pedagogy among university faculty, initially at the department level, and eventually broadening so that ideas can be shared across disciplines. Unfortunately, faculty in higher education are now accustomed to working in disintegrated ways when it comes to teaching. Not only are we uninformed about what our students experience outside of our departments, even our so-called curricula are not constituted by much more than a proximal listing of courses in school catalogues. Innovations cannot survive the innovators if these discussions do not take place. Conversations among the national community of innovators (including chemistry faculty) are happening, and they are stimulated in part by the systemic initiatives programs for curriculum reform. However, the core problems of sustained reform will not be solved unless the behaviors persist after the funding is removed. A series of highly articulated theoretical frameworks for science education need to emerge along with the models for practice.

Progress in science relies on a multifaceted collaborative community of scholarship. We identify undergraduates for their promise in research and then provide those who become faculty with a highly organized structure of professional development that extends beyond graduate school. Something that is only coming to be understood as a question is whether we can think about the development of the scholarship of teaching in ways that mirror what we do in other forms of scholarship. Significant parts of our professional training include learning how to develop ideas and learning what constitutes evidence. In our practice in the physical sciences, assessment and evaluation play an indispensable role in the incremental evolution of our understanding about the physical world. Our use of evidence is culturally embedded. For assessment information to be used in science education in the same way we use it in science, the corresponding development of a cultural context must be provided. Progress in science education is simply another form of progress, and it will rely on developing the same sort of multifaceted collaborative community of scholarship that has worked so well in science.

NISE Forum Trigger Questions

In this section, I have organized short replies to the three trigger questions provided by the organizers of the Forum. In almost all cases, a more complete understanding of my responses can be found by reading the first section.

Learning Assessment and Reform: What role can the assessment of learning play in undergraduate education reform—in departments, among colleagues, and within the disciplines? My assertion is that we still need to do a great deal of work to design and implement a truly morally reflective undergraduate science education. Assessment of student learning, which is indelibly linked to faculty instruction, can provide the kind of information we currently lack about the actual effects of a contemporary science education rather than the imagined effects (and efficacy) that characterize our current practices. Assessment can:

- inspire people to act and plan more thoughtfully
- inspire more public discourse
- promote multidisciplinary conversations and projects
- recognize the need for greater professional development in future faculty, a dramatic and overdue change in the educational culture in science
- provide feedback information for personal improvement by reflective practitioners
- provide feedback to students, to curriculum design, and to decision-making.

As we conduct assessments that are not a part of the science education tradition, there are two important issues that emerge. First, the standards for evaluation of any type of assessment are culturally embedded. A nuclear magnetic resonance (NMR) spectrum of some substance carries an extraordinarily rich amount of assessment information, yet evaluating this information is not self-evident. If nontraditional assessment information is delivered to someone at the same time that they are expected to use it evaluatively, then they will not be able to examine it critically. It is no more likely that an understanding of structural determination can be derived from a single NMR spectrum than understanding the flora of an ecosystem can be derived from a single twig cut from a sapling. Coding and analyzing interviews, collecting and reporting survey data, and so on, all have their own scholarly traditions outside of the physical sciences. A scientist must learn to be literate in the scholarship of these areas in order to make sense of assessment information. The second issue is more provocative. As our confidence in evaluating assessment information increases, there will instantly be situations in which a faculty member can be confidently described as unsuccessful or even noncompliant within a department's educational mission. We must be ready to provide guidance for all evaluative outcomes. Thinking further, this will also create situations in which an individual faculty member in a course will invoke academic freedom in the face of not fulfilling the instructional covenant established by the departmental unit or subunit. I have faced this in my own work, and I cannot provide much insight beyond the need to anticipate this as an inevitable outcome from educational assessment.

As one of the readers of an earlier draft of this manuscript pointed out, I have quite deliberately had little to say about the benefits of assessment for the audience of administrators and funders who are demanding this information. While there may be political benefit for those who are demanding accountability from higher administrators, it is consistent with my argument that these individuals are more interested in using the political results than the particular substantive outcomes. My viewpoint is that assessment information is interesting and important for moving intellectual agendas forward, for contributing to progress. Funders are rarely sources of funds; they are really more like intermediaries who must answer to politicians or boards of directors. So I understand that these results need to be used within the political context, and that there is ultimately direct benefit to the faculty from being able to sustain and increase funding, and so on. But wait . . . "sustain and increase funding"? Is that our

goal? I urge that we must continually reaffirm our commitment to progress toward excellence in and understanding of science education as our primary goal. I consider the benefits from assessment to anything other than science education to be relevant but of secondary importance. Inverting these priorities is like the tail wagging the dog. (And I hope that this statement is not perceived as biting the hand that feeds me!)

Assessment as Argument. How do we best leverage change in these spheres: Can assessment be used to convince colleagues, protect innovators from risk, and build support for educational change? My short answers are "not yet," "not yet," and "that's a good place to start." My longer answers follow.

Can we use this information to convince colleagues? There are false assumptions embedded in this question. Alone, assessment information convinces no one of anything. My father sees an NMR spectrum for what it is: lines of ink on paper. My mother looks at the chemical equations drawn in my publications and sees "little bugs crawling on the page." As I have already pointed out, if the cultural context for evaluation (e-VALUE-ation) is not cultivated, then assessment information just becomes a list of numbers, lines on a graph, or little bugs. Assessment methods and their evaluative contexts need to be understood before the information can possibly influence the beliefs of an individual. Even then, as my earlier examples of poor personal health behaviors should make us realize, there is more to changing beliefs and behaviors than data. In fact, it is a gross misunderstanding about how change takes place in science to think that simply presenting assessment information can cause it to occur. Further assessment conducted outside of the scientific arena is commonly disparaged ("soft") and a frequent target of criticism from within the scientific community.

Can we use assessment to protect innovators? Not yet, for most of these same reasons. Innovators must also include education in their mission if assessments that are not traditional for a given community are going to be used to constitute evidence.

Can assessment build support? Yes, possibly. This points to the strategy implied above: Progress in using educational assessment will rely on how well we bring others into the multidisciplinary culture for which we have built our own respect.

There is a unifying theme here that is familiar in science. As Maxwell is reputed to have pointed out: Science sometimes progresses funeral by funeral. If we have identified new areas of value that we think belong in this enterprise we call science education, then they must become a part of the professional development of the next generation. The attitudes and cultural context that define us are simply what we learned in our own education, after all. If the problem is a core issue, then we must change the development of the next generation at the core. If the value we think is associated with these changes is borne out, then the usual process by which an emergent area ends up encultured will occur.

Effective Dissemination: What forms of dissemination work best in encouraging others to try, to support, or (at minimum) not to obstruct classroom innovation? The same question drives this section: How can innovations survive the innovator? Dissemination occurs along multiple dimensions simultaneously. At one end, innovations that arise from the work of an individual teaching a course need to become owned by a noninnovator who then teaches the course. At the other end, broad systemic recommendations need to be carefully described with appropriate contextually meaningful examples in order for a noninnovator to understand the change at all. Some of the problems associated with group learning are good examples. Because assignments have not been constructed to ensure collaborative interactions, we

repeatedly hear about the simple cooperative model for group work being devised by students: the challenging problems are parsed out to the better group member, and so on down the line, which is a degree of cooperation, and then the collaboration occurs when the group staples their work together.

Although we use the term *dissemination*, we really mean institutionalization and cultural change. Simply getting information into the hands of others efficiently, which is the tradtional view of dissemination, is not an effective agent of cultural change. Again, as argued earlier, information is received within a cultural context. Faculty who are ready to learn new things need to be able to access this information and to participate in a variety of learning opportunities, including written and electronic materials, workshops, and other training methods. The value system in which a faculty learner exists is much more significant than the dissemination system, however, and tangible benefits must accompany the effort to learn the information being disseminated. For students, we have already shifted from "teaching information" to "student learning." Once you shift from a dissemination mode (teaching information) mode to "faculty learning," you open a whole new understanding about innovation. Learning is personal and does not rely on simply receiving information. Learning ultimately happens one person at a time and in many ways. Faculty development strategies cannot focus solely on re-education of the current faculty, or reform will continue to fight the losing battle to "fix" the incomplete education of our new faculty. We need to couple faculty re-education with new strategies to prepare the next generation of science faculty for a more valued career in higher education.

Selected Bibliography (General)

AAHE Teaching Initiative (P. Hutchings, program director; P. Bender, program coordinator) American Association for Higher Education, One Dupont Circle, Suite 360, Washington, DC. Available URL for AAHE at *http://www.ido.gmu.edu/aahe/welcome.html*

Anderson, E. (Ed.). (1993). *Campus use of the teaching portfolio: Twenty-five profiles.* Washington, DC: American Association of Higher Education.

Angelo, T. A. (1991). Introduction and overview: From classroom assessment to classroom research. In T. A. Angelo (Ed.), *Classroom research: Early lessons from success* (New Directions for Teaching and Learning No. 46, pp. 7-15). San Francisco: Jossey-Bass.

Angelo, T. A. (1991). Ten easy pieces: Assessing higher learning in four dimensions. In T. A. Angelo (Ed.), *Classroom research: Early lessons from success* (New Directions for Teaching and Learning No. 46, pp. 17-31). San Francisco: Jossey-Bass.

Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers* (2nd ed.). San Francisco: Jossey-Bass.

Barr, R. B., & Tagg, J. (1995). From teaching to learning. A new paradigm for undergraduate education. *Change*, *27*(6), 13-25.

Boyer, E. L. (1990). *Scholarship reconsidered: Priorities of the professoriate.* Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.

Edgerton, R.; Hutchings, P.; Quinlan, K. *The teaching portfolio: Capturing the scholarship of teaching*. Washington, DC: American Association of Higher Education, 1991.

Ennis, R. H. (1989). "Critical thinking and subject specificity: Clarification and needed research" *Educational Researcher, 18*(8), 4-10.

Ennis, R. H. (1990). "The extent to which critical thinking is subject-specific: Further clarification" *Educational Researcher, 19*(4), 13-16.

Gabel, D. L. (Ed.). (1994). *Handbook of research on science teaching and learning* (pp. 79-81, 113-114). New York: MacMillan.

Glassick, C. E., Huber, M. T., & Maeroff, G. E. (1997) *Scholarship assessed: Evaluation of the professoriate*. San Francisco: Jossey-Bass.

Herron, D. "Chemical education research: Task force report" American Chemical Society Division of Chemical Education: Division of Chemical Education (CHED) Newsletter, Fall, 1996, 23-26.

Hutchings, P. (1993). *Using cases to improve college teaching: A guide to more reflective practice*. Washington, DC: American Association of Higher Education.

Hutchings, P. (Ed.). (1995). *From idea to prototype: The peer review of teaching: A project workbook*. Washington, DC: American Association of Higher Education.

Hutchings, P. (1996). *Making teaching community property: A menu for peer collaboration and peer review*. Washington, DC: American Association of Higher Education.

Lambert, L. M., Tice, S. L., & Featherstone, P. H. (1996). *University teaching: A guide for graduate students*. Syracuse, N.Y.: Syracuse University Press..

Lowman, J. (1995). *Mastering the techniques of teaching* (2nd ed.). San Francisco: Jossey-Bass.

McKeachie, W. J. (1994). *Teaching tips: Strategies, research, and theory for college and university teachers* (9th ed.). Lexington, MA: D. C. Heath.

McPeck, J. E. (1990). Critical thinking and subject specificity: A reply to Ennis. *Educational Researcher, 19*(4), 10-12.

Menges, R. J., & Weimer, M. (Eds.).(1996). *Teaching on solid ground: Using scholarship to improve practice*. San Francisco: Jossey-Bass.

Prichard, K. W., & Sawyer, R. M. (Eds.). (1994). *Handbook of college teaching: Theory and applications*. Westport, CT: Greenwood.

Shulman, L. S. (1993). Teaching as community property. *Change, 25*(6), 6-7.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

van Keulen, H. (1995). *Making sense: Simulation-of-research in organic chemistry education*. Utrecht, The Netherlands: CD-ß Press.

Weimer, M. (1993). *Improving your classroom teaching*. Newbury Park, CA: SAGE Publications.

Selected Bibliography (Author)

Coppola, B. P. (1998). First, do no harm...The (moral) obligation of the faculty. *Proceedings of the 1998 electronic conference on chemical education*. [On-line]. Available: *http://www.wam.umd.edu/~toh/ChemConf98.html*

Coppola, B. P., & Daniels, D. S. (in press). Mea culpa: Formal education and the dis-integrated world. *Science and Education*.

Coppola, B. P., & Daniels, D. S. (1996). Structuring the liberal (arts) education in chemistry. *Chemical Educator*, 1(2).

Coppola, B. P., & Daniels, D. S. (1996). The role of written and verbal expression in learning. Promoting and improving communication skills for students in an undergraduate chemistry program. *Language and learning across the disciplines*, *1*(3), 873-878.

Coppola, B. P., Ege, S. N., & Lawton, R. G. (1997). The University of Michigan undergraduate chemistry curriculum. 2. Instructional strategies and assessment. *J. Chemical Education*, *74*, 84-94.

Ege, S. N., Coppola, B. P., & Lawton, R. G. (1997). The University of Michigan undergraduate chemistry curriculum: Philosophy, curriculum and the nature of changes. *J. Chemical Education, 74*, 74-83.

Kovac, J., & Coppola, B. P. (1997). Universities as moral communities. *Proceedings of the 1997 conference on values in higher education*. [On-line]. Available: *http://funnelweb.utcc.utk.edu/~unistudy/values*

Assessment and the Promotion of Change in Community Colleges

Eileen L. Lewis
Chemistry Department
Cañada College

Today the community colleges are in much the same position as the science, mathematics, engineering, and technology (SMET) education community only a few years ago with respect to assessment—there is very little! One stimulus for improving both the quantity and quality of assessment in the SMET education community has been the NSF's requirement that all proposals include plans for meaningful evaluations. However, there has been virtually no evaluation of what we call traditional curricula in SMET. The outcomes from traditional curricula have simply become the default standard to which we compare reform. We have no real evidence to show that existing curricula or methods used in SMET education are particularly effective or desirable, or how the effect of those curricula and methods vary over populations.

An example may be useful in illustrating how much our thinking has changed over the last few years. In the mid-1980s I was involved in a project whose goal was to develop new ways for each of the three colleges in my district to teach our Preparatory Chemistry Class (a class for students with no previous chemistry to prepare them to be successful in General Chemistry). As we looked at a variety of chemistry reform projects in both two- and four-year institutions across the country, assessment of those reform projects was conspicuously absent. (I should note that we did not notice the absence or seem to expect such assessment.) What constituted evidence were anecdotes about how students really liked the new curriculum, how faculty liked teaching using these new methods (primarily variations of a Personalized Systems of Instruction (PSI) and Mastery Learning), how enrollment was up in the one or two semesters the project had been in use, or how cool it was to use microfiche. We chose to implement a reform project that had been developed by a very successful community college textbook author and adapted by another community college. Our decision was based on the way the content of the course was parsed, the way it used active and mastery learning, and the comments we heard from the few students and faculty we interviewed at the two campuses.

A more recent larger-scale example can be found in the Calculus Reform Projects funded by NSF. These broad-scale projects that resulted in much systemic reform had no assessment or evaluation component attached to them, nor was evaluation a required component for NSF funding. It is easy to see how recent the concept of meaningful assessment and evaluation in SMET reform projects is at the college level.

A third, even more recent illustration of the assumption that evaluation is either unnecessary or optional occurred at a presentation at the 1996 ACS meeting in New Orleans. The speaker was a chemistry professor at a large well-known Midwestern university. After presenting information on his reforms in a particular course, he proceeded to tell the audience how effective those reforms were. No one in the audience questioned his comments. After a few minutes, I asked what methods he had used to evaluate the reforms and what evidence supported his assertions. He waved his hand in the air with a somewhat dismissive gesture and replied, "Oh, I haven't done any of that education stuff yet."

I do not mean to be critical. It is hard to be an expert in more than one field, and well-

designed studies require both time and knowledge. But I would argue that neither he nor any of us would make a similar assertion about some scientific phenomenon without sufficient evidence, yet we have all made similar assertions about our efforts at reform. While the NSF is promoting and supporting meaningful evaluation by requiring it as a part of every SMET proposal funded, evaluation must reflect curricular changes. We simply cannot use old measures with changing curriculum.

Students are, by necessity, very cognitively efficient. They will adapt their learning methods and content to our style of assessment. We say we care about deep understanding and the ability to apply that understanding in a variety of settings, yet our exams often measure what is easy to measure—recall and algorithmic problem solving. Why are we then surprised when students focus on this type of learning. It has also been my experience that on an exam students often appear to know more than they actually do. As an example, I was engaged in research to characterize students' thinking and learning processes over the course of a semester-long class using microcomputer-based learning (MBL). As part of the research, I interviewed students at regular intervals and discovered that performance on exams was not a real measure of understanding. Superficial knowledge went a long way on the exams. Students could often use terms in approximately the correct ways and contexts. They also were able to solve problems without understanding their meaning. It was not uncommon to find students discussing insulators and conductors, seemingly coherently, describing metals as conductors and materials like wood as an insulator. However, upon deeper probing, it became clear that they thought conductors felt cool in cold environments and warm in hot environments because they have the ability to attract, absorb, or hold cold/heat (Lewis & Linn, 1994). This is a simple example, but I assure you that, if you want a surprise, interview your students and probe the depth of their understanding of a concept you think they understand. Another telling experience comes if you ask a student some question about the meaning of an answer they have just calculated. They look at their work and restate their numerical answer, often without units. If we then ask, "But what does it mean? How does what you calculated relate to the question?" They look blank and act as though no one had ever asked that before.

The point is, if we want student to go beyond science facts to the methods and processes of science, our assessments must reelect this aim. We say we care about meaningful understanding, about the application of concepts to new contexts, and about knowing how to learn, but until our assessment match those goals, students will have other cognitive goals.

Thus, it is inappropriate to use our current or standard assessments with innovative curricula. A typical goal of innovative curricula may be a deeper understanding of concepts through the exploration of those concepts in complex real-world settings. Additional goals may relate to learning the methods and processes of science, such as an improved ability to analyze data and to use evidence to support or refute a hypothesis. Assuming the goals of innovative curricula are attained, why would students do any better than their traditional counterparts on assessments that focus on algorithmic problem solving and recall? The assessments we use must match course content. The old horse-race model for evaluation will not distinguish conceptual differences between participants.

The Current Role of Assessment In Community Colleges

With respect to the community colleges today, a common form of assessment is to track changes over time in the number of Weekly Student Contact Hours (WSCH), i.e., the total number of student hours that each faculty member spends per week. Thus, if 30

students were in a class that met six hours per week, that would be a WSCH of 180. Another measure is to determine the ratio of WSCH per full-time faculty equivalent (FTE) or WSCH/FTE, i.e., weekly student contact hours/full time equivalent faculty member across classes, instructors, or subject areas. This value system is based on sheer numbers and favors classes with large enrollments that reflect the funding methods for community colleges, which is based on ADA (average daily attendance). Most community colleges also track retention and persistence rates or the number of transfers to local universities. However, assessments of student learning, conceptual change, or performance in subsequent traditional or reform classes would have to be done by faculty members who teach fifteen semester units, do all their own grading, and are simply not trained in assessment or evaluation processes. As a result, unless faculty have substantial internal or external funding for a project, meaningful evaluation, in terms of student learning gains is unlikely to occur.

One of the most effective ways for evaluation to truly impact the community colleges is through papers and reports that synthesize and disseminate research findings from science education and research in cognition. Faculty need to see research findings on passive versus active learning, good classroom practices, new pedagogues, and the use of technology. Such findings can help convince colleagues, department chairs, and deans that certain kinds of changes in the way we design and teach our classes will improve student skills, understanding, and performance. Leadership is also provided by reports such as the 1996 Advisory Committee (ACEHR) report to the National Science Foundation, *Shaping the Future: New Expectations for Undergraduate Education in Science, Mathematics, Engineering, and Technology*. This report proposes that faculty should actively engage students, assist them to learn not only science facts but also the methods and processes of research, and model what scientists and engineers do. They should give students the tools to make informed judgments about technical matters, enhance their communication skills, and enable them to work in teams to solve complex problems. The impact of such urging from so important a national funding agency enables and legitimates the classroom innovator.

Faculty not only need research findings to support their reform efforts, but they need materials and experiments that help them implement new curricula. As we have discovered in testing a new curriculum for a second semester general chemistry course, faculty need to understand the intent of authors when concepts are embedded in entirely new contexts. Again, this is particularly true of community college faculty who teach fifteen semester units and have limited time for development activities. In addition to written materials, faculty benefit from workshops that help them see and understand transformations in our understanding of teaching and learning. Workshops also can be of great benefit in generating excitement to try new ideas as well as a resource in convincing reluctant colleagues. Seminars and workshops are also critical to sharing information about new reform curricula—how to teach them, and what assessments have shown about student outcomes—and are a very effective way to have both willing and reluctant colleagues experience new content and pedagogues. Individuals often say they welcome change until it comes to their actually making it. Thus, the assessment information and support received in workshops is vital to educational reform efforts.

Another consideration in any reform effort in the community colleges is the faculty and administrative concern with student preparation for successful transfer to four-year institutions. We are sensitive to both what is taught and how it is taught at those four-year institutions. There is reluctance to make major methodological or curricular changes--such as eliminating concepts,

focusing on depth versus breadth, or using new pedagogues--if they represent too great a departure from what is perceived to be practiced at four-year institutions. This reluctance is waived, however, where such changes are broadly supported by the SMET education community. There is also no question that legitimacy is provided by whatever major universities, NSF, and foundations do, and what they are perceived to support.

Case Studies Of Reform

A young man with some part-time teaching experience whom we shall call Mark was hired in a tenure track position in mathematics. He knew how students learn from his education classes in graduate school and was committed to reform because he felt his classes were not working. Mark was frustrated by students' ability to do symbolic manipulations with little understanding of the concepts behind those manipulations. While he reported that the stated goals of the existing curriculum were to help students develop skills, the only outcome appeared to be what he called "symbol pushing." One aspect of symbol pushing was that students regularly mimicked processes in an inappropriate context. One of his goals was to have students understand concepts well enough to be able to recognize them at work in other contexts, e.g., engineering or chemistry. His frustration coincided with the availability of new textbooks and materials that incorporated mathematics reform efforts. He freely admitted that he was not sure he would have put forth all the effort required to rewrite whole curricula.

When he proposed adoption of a new calculus reform textbook that used graphing calculators and graphical representations, there was some passive opposition from some senior faculty, particularly those who planned retirement within a few years. One faculty member who opposed the new book simply stopped teaching calculus to avoid using the new text. Some concern was expressed about the lack of polish and frequency of errors in any new textbook. As a concession there was general agreement that preliminary editions of books would not be used.

Mark credited several factors for his success in getting calculus reform efforts going and then moving that reform down the curriculum into algebra. He did not suggest implementing reform when he first arrived, partially because he wanted to become a part of the community, and partially because he was teaching several courses for the first time. He stated that it was easier to do things the way they had always been done, and that often the best way to teach a new class is to begin by teaching it in the traditional way to find out what does and doesn't work.

Mark's advice to new faculty is to build friendships, ask advice from peers, volunteer for professional activities when the need arises, and share your ideas with colleagues progressively. Of course, all this can be called becoming collegial in a new environment.

Another source of support for Mark's changes came from published literature. While he had seldom read the original studies, publishers of reform books sent out synopses of those studies that he was able to reference in building support for a reform curriculum. Additionally, the mathematics professional societies held conferences that discussed reform efforts and outcomes and workshops on how to teach using the reformed curricula. He also credited colleagues and a supportive dean who were open to change making his reform activities possible. Mark was able to progressively introduce changes into his department and was successful in the tenure process.

While all the calculus classes in this instance adopted the same new reform textbook, the suggested reform books were not uniformly adopted in algebra. Some faculty continued to use their own books and methods. In community colleges, more discretion may be allowed to individual faculty in the choice of textbooks for their

classes than in four-year institutions. However, the degree of that discretion with respect to textbooks varies greatly from campus to campus.

With respect to assessment, Mark admits that he set up no studies and collected no systematic data. His information is entirely anecdotal and centers around the accounts of students in subsequent classes or who return to tell their success stories. One example is that of a returning student who had transferred to Stanford University and was taking an economics class. He proudly reported that his economics professor used all sorts of equations describing economic phenomena, and while other students did not know what the equations meant, he did.

Mark told a parallel story of an old friend who also began a tenure track position at a sister school within Mark's community college district. His friend began by pushing immediately for reform and found himself at odds with his colleagues. He too provided little external data to support his position. One of the many criticisms leveled by his colleagues was that his use of group work resulted in covering less material. During the tenure review process the curriculum he had created was used against him rather than for him. Peer evaluations were poor, and he was only granted tenure because of the efforts of a very supportive dean. Based on other stories gathered from faculty and deans, I suspect that this individual was lucky to have gained tenure at all. Regardless of tenure outcomes, it is undesirable from everyone's perspective that a creative faculty member be alienated and isolated from colleagues.

In interviewing deans and faculty for this paper I was occasionally warned that innovation was thought to be risky. For example, setting clear performance objectives was risky since they may not initially be met when using new materials and methods. In some community colleges, the administration does not view innovative faculty favorably because the administration wants to control the activities and direction of the departments and of the college overall. Another risk comes from established faculty who may resent a young faculty member who comes in with ideas and skills, e.g., the use of technology, that are favored by the administration. I like to think that the situation where innovation is risky is rare, but evidence suggests that this is an issue that should be considered. Certainly a primary consideration in any reform activity is to know your colleagues and know your institution. I was also told of conflicts that have arisen between the faculty union and administrators over innovations proposed by faculty.

Community colleges may be unique in that there may be two different groups representing the faculty. The Academic Senate generally focuses on the academic and professional concerns of the faculty. Since most community colleges have collective bargaining, they are also represented in these processes by a union concerned with working conditions. There are times when these two representative groups conflict. Several years ago, the State Academic Senate for California Community Colleges proposed a more comprehensive and meaningful faculty evaluation process. This process was adopted by the state Chancellor's Office and general standards were set to be used throughout the state. While in my district the union, senate, and administration worked together to define the process, there was a built in conflict between "meaningful evaluation," with its attendant consequences and the union's role of protecting an individual faculty member's rights. Depending upon local circumstances, this conflict has both aided innovative faculty and pitted them against their unsupportive peers.

Although these observations are drawn from a limited number of personal accounts rather than from systematically gathered data, some clear messages emerge that ring true with our experience of working in institutions. If your goal is reform and you are an established and respected faculty member, your task will be easier. If you are

new to an institution, you have a greater chance of succeeding in your classroom reforms if you take the time to know your colleagues, their background experience and values, as well as the culture of your institution. It is also crucial that you take the time to help colleagues truly understand what you are proposing. Very much in keeping with Strike and Posner's requirements for conceptual change that is both intelligible and plausible, fellow faculty must both understand your ideas and not find those ideas too far different from their own. It is also a great asset if a college has some discontent with current practices (Strike & Posner, 1985, 1992). As faculty seeking innovation, you should enlist support from colleagues who are open to reform ideas, share your ideas and struggles with colleagues, enlist support from your department chair and/or dean, share reference materials that support your proposed ideas, attend conferences with colleagues, and get them involved in reform efforts. Faculty can and do change, but you may need to convince them that change is worth the effort required.

Classroom innovators in community colleges have an advantage over those in four-year institutions in that teaching is the foremost concern of faculty and administration and the major criterion in all faculty evaluation processes. With this focus, it is easier to start conversations around questions such as, "What should we be teaching? What is the best way to help students gain the knowledge and skills they need to be successful in the world today?" The very act of getting faculty to reflect on their own goals, consider how to best achieve those goals, and how to assesses that achievement may be a profound form of faculty development. It is hard to return to old ways of thinking about learning and teaching once you have begun to this process and experience students actively engaged and talking about ideas in science.

We can also benefit from viewing reform efforts within larger context of Rogers' *Diffusion of innovations* (1995). Rogers writes about five archetypes active within transformation processes: the innovator, the early adopter, early majority, cynics, and laggards. Perhaps considering the effect of meaningful assessments on each of these archetypes will help us frame compelling arguments that will be effective for a wider array of institutional environments.

References

National Science Foundation. (1996). *Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology* (NSF 96-139). Washington, DC: U.S. Government Printing Office.

Lewis, E. L., & Linn, M. C. (1994). Heat energy and temperature concepts of adolescents, naïve adults, and experts: Implications for curricular improvements. *Journal of Research in Science Teaching, 31*(6), 657-677.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A.(1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education 66*(2), 211-227.

Rogers, E. M. (1995). *Diffusion of innovations* (4th ed.). New York: Free Press.

Strike, K. A., & Posner, G. J. (1985). A conceptual change view of learning and understanding. In L. H. T. West & A. L. Pines (Eds.), *Cognitive structure and conceptual change*. New York: Academic Press.

Strike, K. A. and Posner, G. J. (1992). A revisionist theory of conceptual change. In R. Duschl, & R. Hamilton (Eds.), *Philosophy of science, cognitive science, and educational theory and practice*. Albany, NY: SUNY Press.

Assessing and Evaluating the Evaluation Tool: The Standardized Test

Richard Tapia
Noah Harding Professor
Department of Computational and Applied Mathematics
Rice University

The misuse of standardized tests at selective and even not-so-selective institutions prevents the nation from tapping into a large part of its human resources' creativity and leadership. We are significantly retarding the processes of change and reform that have been accepted as critical to maintaining our national health. For decades now, we have let the traditional beliefs of the ruling class dictate the policy for change and reform in testing, and consequently we have ended up with little or no reform. My purpose in writing this essay is to push for rigorous study of standardized tests' traditional use. It is imperative that we collect data, evaluate and assess, and use these findings as the impetus for change and reform. While we often allude to such studies, they are invariably incomplete, anecdotal, or nonrigorous. Hence, there can be no effective dissemination or buy-in on the part of our colleagues, administrators, and national educational policymakers. This problem is not restricted to underrepresented minorities; although, I quickly add, as identifiable groups, they are hurt the most. Indeed there is good correlation between the misuse of standardized tests and underrepresentation. Bluntly put, the misuse of standardized tests is the underrepresented minority's worst enemy.

Let us suggest an effective way of using SAT scores that we shall refer to as a "threshold approach." In most selective universities, admissions people are going to look at the higher end of the test spectrum, say 1300 and above, and then try to make decisions from that group. I maintain that, in terms of SAT score alone, we can not make meaningful distinctions in terms of real success between members of the group consisting of individuals who have scored, say 1050, and above. Moreover there are significantly many individuals with SAT scores between 1050 and 1300 who, in a real sense, will be equally or more successful than most individuals with scores above 1300. So, 1050 is our threshold value. All with scores above 1050 are deemed acceptable, and other factors should be used to differentiate among the members of the acceptable group. All other factors being equal, I have no problem with breaking the ties with SAT scores. On the other hand, experience has taught me that it is unlikely that individuals with scores below, say 850, will succeed at Rice University, so we shouldn't accept them into Rice. Now what can we say about the group of individuals with scores between 850 and 1050? Well, we need to look very closely at them and decide whether they should be put in the reject class, the acceptable class, or some other class that would require additional information and study.

Rice University has been quite successful at implementing diversity in its undergraduate population. The threshold system deserves much credit for this success. The Rice Guidelines for Admission and Financial Aid included in section 3 of this paper strongly allude to a threshold approach to the use of SAT scores in the undergraduate admission process. On average, Rice University underrepresented minority students have substantially lower SAT scores than does the student population at large. However, they are on par with their nonminority counterparts in terms of retention rate and grade point average. They bring in more than their share of awards and

admissions to prestigious graduate and professional schools.

I realize that the parameters used in my presentation of the threshold approach to the SAT score are somewhat arbitrary. In a real situation they would have to be fuzzy numbers. However, it is really more the concept that I want to discuss in this section.

I would like to preface my remaining comments with three anecdotes. The stories are true; they really happened. The first concerns a Mexican American male, the second a white male, and the third an African American female.

Pedro was born in the barrios of San Antonio, Texas. He was proud of his Mexican-American and barrio heritage; in fact he was so proud that he had no problem referring to himself as Chicano. Some at Rice University felt that he was a "Barrio elitist." He gave them the feeling that, if you were not from the barrio, you really did not have a handle on life and did not know what life was all about. From his traditional family Pedro learned to be respectful and considerate to others. From the barrio he learned a sense of survival and toughness. He could be sensitive and he could be tough as the situation required. Pedro possessed excellent mathematical and scientific talent. He was the star of his local barrio school. Of course no one from his barrio school had ever gone off to a selective college like Rice; in fact few had gone anywhere except community college. The combination of academic success, inner-city survival, and pride of his heritage gave Pedro considerable inner confidence and self-esteem. He had learned not only how to survive, but also the excellent attribute of never quitting or giving up.

While Pedro and his family had never heard about Rice University, one of his counselors knew about Rice. He advised him to apply to Rice. Pedro had excellent grades and letters of recommendation. His SAT score was 400 points below the Rice average of 1410. However, Rice University was making strong efforts to improve diversity, and we were experimenting with what I call the threshold approach to standardized tests. More will be said about this later. Suffice it to say for the purpose of this story, the threshold approach essentially sets a threshold score for acceptability and deems all above that score as equivalent with respect to SAT score. Decisions are then made on members of this equivalence class by considering other factors. So, Pedro was accepted at Rice. In previous years when the threshold policy was not in play he would have been quickly rejected. Pedro found Rice very demanding and very challenging. He received several C grades. He thought of leaving. But he was not a quitter and he stayed. When I met Pedro in his junior year at Rice he was president of an active Hispanic organization at Rice. He took a class from me in mathematics. In class it was clear that he had excellent scientific talent. I found him to be exceptionally creative. He was not the best student in the traditional sense, but he was very good, and no one seemed to me to have better potential for graduate school. So, I asked him if he planned on attending graduate school. He replied that he had received several grades of C early on in his Rice career. I told him that the grades alone would not preclude his acceptance, especially since they were early in his career and he was doing so well now. He was very excited, took the GRE and applied to several good schools; Stanford, Berkeley, the University of Texas, and Texas A&M University. I wrote him a very strong letter emphasizing that he was not only an excellent student, but one of the more creative students that I had taught at Rice in 25 years. School by school rejected him saying that his GRE scores were too low. I then pushed strongly for his acceptance at Rice and was successful, again because we use a form of the threshold approach in some of our departments' graduate programs. He breezed through a thesis masters degree. As before our faculty gained a high respect for his talent and creativity. He had an opportunity to work for Texas Instruments here in Houston while finishing up the Ph.D. degree. At Texas Instruments he has been a star. Recently his supervisor asked me if we had any more like Pedro; he said that he would hire as many as we had. Pedro will finish his doctorate this year with an excellent dissertation.

The moral to this story is obvious. If I had not played a major role Pedro would not have realized his full potential and his leadership would have been lost to the scientific community. He would not have had the chance to become the leader that he has become in both the industrial and academic communities. He would have been cut down by his GRE score. Misuse of the standardized test would have claimed yet one more victim.

Jim is a white male who grew up in a small Texas town. He told me that he always knew that he was smart, but things were not in proper alignment and he dropped out of school in the ninth grade. He eventually moved to the Houston area and decided to obtain his GED from a local community college. In community college he was an absolute star. He was directed to Rice University for his undergraduate education. Rice is extremely selective and rarely pulls from the community college population. However, it is to our credit that we accepted him. As was explained in the previous story about Pedro, Rice does not put overly excessive weight on the use of standardized tests at the undergraduate level (however, wait for my third story). At Rice, Jim took several advanced mathematics classes from me. One of the classes is essentially a graduate course, and Jim was the star of the class as an undergraduate. He clearly was one of the more mathematically creative students that I have taught in all my 27 years of teaching. We talked often and I encouraged him to apply to good graduate schools. I wrote him a very strong letter. Not long ago, Jim appeared in my office very somber and distraught. He confided in me that he had been rejected at Berkeley, Cornell, and Stanford. He had been accepted at one good state school and at Cambridge and Oxford in England. He desperately wanted to know what had gone wrong. I asked him about his GPA. He said, "I will be graduating from Rice with an A+ average and the distinction of summa cum laude. Moreover, I did it in mathematics, one of Rice's most challenging majors, and I did it in three years coming from a community college high school GED." I then asked him about his reference letters. He quickly replied that all who wrote were professors from classes where he was at the very top of the class and they had told him that their letters were very strong. Finally, I said, "Tell me about your GRE scores." He answered that in two of the three categories he had done very well, but in one category he was only in the 75th percentile. I replied, "That's it." He said, "How can that possibly be?" I repeated my reply, and we had a much needed conversation.

Jim was a victim of the misuse of a standardized test. Yet he was a white male, a straight A student at a very demanding school, and one of the most intelligent and creative individuals that I have ever had the pleasure of teaching.

Sandra is African American and was born and raised in Houston, Texas. She was an excellent student in high school and received a full scholarship to study at a university in the northeast, well-known for its excellent engineering programs. Upon graduation she applied to graduate school at Rice University, to one of our "better" engineering departments. She felt that it would be nice to return to Texas, and Rice had a fine reputation. She applied, had not been accepted, and was visiting Rice. I was asked if I would be willing to talk to her. I replied that I would be happy to talk to her. She was brought to my office by a faculty member that I have respect for professionally. I spent considerable time with Sandra, in my estimation she was a potential star. She had an A average from an excellent school, was very mature and focused, had overcome serious obstacles, knew what she wanted and why, and in general was most impressive. I expected the departmental representative to proudly tell me that they were going to accept her and support her. Instead I was asked whether she could be considered for tuition and support under a diversity program that I administer. I asked, "Why? She is an outstanding applicant." I was told that they actually had several applicants that they yet had not decided to accept who were definitely superior to Sandra. I asked in what way they were superior. I was told that the applicant at hand was only in the 89th percentile on one part of the GRE test and the other applicants

(all foreign) had GRE scores in the 92nd or 93rd percentile. Hence, the department felt that it could not pass them over for an "inferior" applicant. It was difficult to contain myself. They were sincere, and of course extremely naive. I chose to support Sandra and lost even more respect for that particular department. She may or may not come to Rice.

As was the case with Pedro, Sandra would not have been accepted if I had not intervened. Our evaluation system is flawed, and it is not going to be saved by waiting for these interventions from outside.

Today universities are looking for individuals with a broader range of attributes. However, standardized tests do nothing to identify most of these attributes. I firmly believe that members of underrepresented groups, by the very nature of being a member of such a group, have learned skills and have developed sensitivities and understandings that would fall into this broad range. For example, in research university environments we talk about the needs for nurturing, mentoring, more effective teaching, a better understanding of the whole student, and outreach to broader communities. Members of our underrepresented groups are prepared to contribute in these directions. However, to a very large extent, these individuals do not have an opportunity to demonstrate this creativity and leadership skill because of traditional assessment barriers. These barriers are not outright discrimination; they are much more subtle. On the surface they look like reasonable measurements of necessary prerequisites or skills. However, they are strongly biased toward the precocious attainment of various pieces of information and knowledge. Potentials for success—creativity, the ability to guide and lead, the ability to adapt to a new environment and bring needed understanding from another environment—are not measured. We do not know how to do this. Moreover, our basic leadership is not totally unhappy with the traditional process, since, after all, their careers were spawned by the process in place; there must be some real good in this current traditional version. While I am basically criticizing the use of standardized tests in undergraduate and graduate admission processes, it is a straightforward matter to extend my criticism to hiring policies, promotion policies, and selection procedures for prestigious fellowships, grants, and other professional rewards. Moreover, while I find it easy to argue in terms of the effect on members of underrepresented groups, I certainly do not wish to imply that these statements and concerns are restricted to them. We are in danger of locally restricting participation that would globally be of value to our national agenda. Local values and global values are usually at odds indeed, often without our being aware of this conflict. The department does not worry about the division, and the division does not worry about the whole university.

A couple of years ago, I served on a committee to review education and human resource development activities of the National Science Foundation. The committee was quite taken aback to find that essentially all the winners of the prestigious fellowship awards were nonminority males who had demonstrated an affinity for science by the time that they were 10 years old or so. The winners were very impressive and undoubtedly very precocious. It was easy for us to feel that the door had been closed on those who were not extremely precocious. Of course our concern was whether this is in the best interest of the nation. We drafted the following statement as a part of our recommendations:

> Committee to Review Education and Human Resource Development Activities of the National Science Foundation (March 1996)
>
> The committee feels that the implementation of the current evaluation criteria concerning the quality of the applicants overemphasizes the "focused prodigy" profile. Since it is

impossible to disentangle productivity due to privilege from productivity due to talent, reviewers and panelists generally fall back on this profile as a means of evaluating candidates, even though it may not be a good predictor of scientific creativity and success. This emphasis works against some candidates (most often women and underrepresented minorities) who may not have been interested in mathematics or science as a young child, but who develop rapidly and demonstrate great creativity once the interest is manifest.

Our Addiction to the Use of Test Scores

For some not well-understood reason university admission committees demonstrate an addiction to the use of one-dimensional qualifiers like the SAT and the GRE test scores in the admission evaluation process. There seems to be a belief that all students can be well-ordered; hence we should try to order all students well. Clearly no two students are the same; therefore we should be able to come up with some measurement that will differentiate. In mathematics we know that it is not possible to well-order quantities that display many components of value. We know in the admissions process that we value many student attributes, yet we fall back on the one-dimensional standardized test. It does get us out of our dilemma, and perhaps this is the most valued aspect of the test. It is simple to use, and it is readily available. It allows us to differentiate with some feeling of security between any two students. It gives us a simple tool. We know that this simple tool can't be perfect, but no one really knows how good or how bad it is; hence for convenience let's use it until someone demonstrates that it is totally flawed. But it works, we get good students; they perform well and succeed, furthermore it is not at all clear that it ignores truly qualified students. So, it can't be that flawed.

People are multidimensional. Science is not only multidimensional, but it benefits from multidimensional approaches. Evaluation from standardized tests places all the weight in one dimension. What about the other dimensions, are they not important? Boldness and creativity play critical roles in research activity; yet, at best, they have a weak correlation to scores on standardized tests. The rub is that, once we concede that the problem is multidimensional, then we don't know what to do. The evaluation process becomes extremely difficult. There is another deficiency in the way we evaluate. We define success in a manner that may not be meaningful. For example, the MCAT score may be a fair predictor of success in medical school. However, success in medical school may not correlate well with success as an effective physician. What is real success? From this point of view, the MCAT is an absolutely hopeless evaluation tool. We value what we measure, because measuring what we value is simply too hard to do.

Test Scores and a Perceived Lowering of Standards

It is interesting that a significant part of our population equates lowering of standards, or an inferior applicant, with scores on a standardized test. This was the essence of the infamous Baake decision in California and Hopwood decision in Texas. So-called inferior minority students were accepted over the named plaintiffs of Baake and Hopwood. Why were these minority students inferior and less capable? Solely because they had lower LSAT scores. We equate lower scores at all levels with lower standards. I have seen highly intelligent colleagues argue the merits of a 93rd percentile GRE score over that of a 90th percentile score (recall our anecdote concerning Sandra). The individual with the lower score was rejected in favor of the one with the higher score with no doubt whatsoever that the process was fair. We have learned to put great value on what we measure and have forgotten to ask whether this measure is flawed concerning what we value. These tests are far from God given. We must evaluate the evaluation criteria. We are

so naive as a nation that we spend considerable time, money, energy, and rational comment on a criterion that is blindly accepted. Here we need to play philosopher more and mathematician less. We must question the validity of the axioms and not just follow the implications of these axioms. However breaking away from traditional use of these standardized tests will be nearly impossible. We need to assess, evaluate their effectiveness, and then use them in the appropriate fashion.

Jesse Shapiro, last year's valedictorian at New York's prestigious Stuyvesant High School, stated in his valedictorian address,

> Nothing could be fairer than a simple multiple-choice exam. It leaves no room for political patronage, racial bias, or other discrimination. Unless New York wants talented-blind admissions, it should keep testing.

Let's stop and reflect on Shapiro's comments. It is easy to be fair. But is being fair the complete picture? "Not fair" would lead us to believe that the process should be questioned. However, a fair process may also have some serious deficiencies. Years ago, my son raced BMX bicycles. I questioned a lane selection process that was being used. I told the officials that it could be greatly improved. They told me that it was fine as it was, because it was fair; each rider had the same chance of getting any particular scenario. I asked them to consider the following hypothetical scenario. Two riders go to the gate. One rider will have to start backwards (rear wheel on the gate) and a coin will be flipped to see which rider has to start backwards. I told them such a procedure was fair in their sense, but was far from optimal. Indeed, I knew exactly how to improve it. They conceded my point, and we eventually introduced a new lane selection system nationwide. An additional point is that multiple-choice tests may be fair, but they rarely test what you want to test.

In the opposite direction from Shapiro's comments, we quote from the Rice University admissions guidelines:

> First, we seek students, both undergraduates and graduates, of keen intellect who will benefit from the Rice experience. Our admissions process employs many different means to identify these qualities in applicants. History shows that no single gauge can adequately predict a student's preparedness for a successful career at Rice. For example, we are cautious in the use of standardized test scores to assess student preparedness and potential. In making a decision to admit or award financial aid, we are careful not to ascribe too much value to any single metric, such as rank in class, grade-point average, the Standard Achievement Test or Graduate Record Exam.
>
> Rice University seeks to create on its campus a rich learning environment in which all students will meet individuals whose life-experiences and world-views differ significantly from their own. We believe that an educated person is one who is at home in many different environments, at ease among people from many different cultures, and willing to test his or her views against those of others. Moreover, we recognize that in this or any university, learning about the world we live in is not by any means limited to the structured interaction between faculty and students in the classroom, but also occurs through informal dialogue between students outside the classroom.
>
> Rice places a premium on recruitment of students who have distinguished themselves through initiatives that build bridges between different cultural, racial, and ethnic groups. In so doing, we endeavor to craft a residential community that fosters creative, intercultural interactions between students a place where prejudices of all sorts are confronted squarely and dispelled.
>
> Our admissions process precludes any quick formula for admitting a given applicant or for giving preference to one particular set of qualifications without reference to the class as a whole. An inevitable consequence of this

approach is that some otherwise deserving and well-qualified students will not be admitted to Rice. By selecting a wide range of matriculants of all types, the admissions process seeks to enrich the learning environment at Rice, and thus increase the value of a Rice education for all students.

(Rice University, Guidelines for Admissions and Financial Aid)

What Is in a Test Score?

In a complex world, be leery of easily quantifiable criteria. In undergraduate admissions there is evidence to believe that SAT scores have some meaningful correlation with first year grades. The question that must be asked here is whether grades are an end in themselves, or just an implied, and perhaps ineffective, predictor of some other meaningful property. In graduate school, grades are never the dominant issue. There is more concern for creativity and an ability to perform new and independent investigations that lead to new theory. Does the GRE score measure this ability or, even more to the point, can it be used to predict success? Bowen and Rudenstine, in their well-known text *In Pursuit of the Ph.D.*, argue that traditional evaluation criteria employed by today's graduate admissions committees do not do a good job of predicting success.

In my years of experience at Rice University on both undergraduate and graduate admissions committees, I have seen many diverse students come through our doors with varying degrees of success and varying levels of scores on standardized tests. I am prepared to say that students with very low test scores will not succeed at Rice. The SAT and the GRE tests are effective predictors of failure for those who score very low. I am not prepared to say that students with high test scores will succeed, particularly in graduate education. I have seen students accepted into our graduate program with excellent undergraduate grades coupled with excellent GRE scores, and yet, from the very beginning, they displayed other attributes, including a perceived lack of creativity, that made me seriously question their ability to succeed in our program. Moreover, they did not succeed. On the other hand we have accepted students with only reasonable GRE scores who were quite successful.

What Is Success for Today's Graduate Student?

In traditional mathematics graduate programs, we have screened and evaluated our students with the implied objective of looking for the next Gauss, or Newton, or Einstein. The loss of an individual who could not measure up was not really a loss according to the accepted objective. Well, perhaps they didn't have to measure up to this extent, but they should be able to be successful faculty at any good research institution in the country. However, only a miniscule number of today's Ph.D. recipients are able to obtain faculty positions at research universities. The vast majority obtain employment in a host of different areas. Many are employed by industry, government, the business world, or nonresearch teaching colleges. Things have changed; the job market has taken on a completely new look. Yet we evaluate, select, train, and educate according to our out-of-date objective. Today's student needs different skills and different training. But our more traditional departments don't change. They continue business as usual. A major point here is that, without a change in the evaluation and assessment procedures, we are undoubtedly excluding students who could excel in the job market, and the new world, and on many occasions producing students who do not fit well into the new job market. We should also realize that college degrees today play the role that high school diplomas played years ago, union cards for fairly nontechnical jobs, e.g., sales. Universities are playing different and

broader roles, yet our admissions policies don't reflect these changes.

The department that I represent at Rice, the Department of Computational and Applied Mathematics, is a world-class department in the area of computational and applied mathematics. Our graduate student population is over 50% women and about 35% underrepresented minorities. These representation figures are unique within the collection of mathematical sciences departments of research universities. Retention through Ph.D. degree is essentially the same for our women and minority students as it is for all students. Our minority students are very qualified. They come from various schools with excellent grades, and all have previous successes. However, as a group their GRE scores are somewhat lower than many of our other students. Their scores are not low, but they also are not what would be required at most other selective universities. Our women and minority students succeed in their graduate work and go off to successful careers. Most go into industry or government research labs. Several of our minority students have demonstrated strong national leadership. We have learned not to put excessive emphasis on the GRE score in the application evaluation process for all students.

While we are on the topic of graduate student diversity, I would like to relate one of my more satisfying teaching experiences. Two years ago I taught a graduate course in mathematical optimization theory. There were 24 students in the class and 12 were members of underrepresented groups, in this case African American and Mexican American. Some of the minority students sat in the front, some in the back, some asked good questions, some didn't ask any questions, some asked questions that did not need to be asked, some did well on the exams, some did not. The class atmosphere was one of genuine interaction. At the end of the semester the majority students had learned a very strong lesson; the minority students were just like them, in that on essentially any professional issue they represented the complete spectrum and could not be stereotyped.

In summary, further study is desperately needed on the way we use SAT and GRE standardized tests in admissions practices. I have put forward a new model that could take us a step in the right direction. This model is not intended as an ultimate solution, but as a way to demonstrate that with some effort we can improve the situation.

Panel 3:
The Role of Evaluation in Institutional and National Policy and Practice

Assessment, Evaluation and Accreditation: Are We on the Same Wave Length? Or How Does One Provide Linkages for Systemic Change?

Jack Bristol
Professor Emeritus
The University of Texas at El Paso

Having spent 27 years as a faculty member and administrator at The University of Texas at El Paso (UTEP), a major urban state-supported institution, I find myself on the verge of understanding a loosely linked and even, sometimes, disconnected series of systems that comprise the higher education world. As a biologist and the author of journal articles on host-parasite physiology, I have come to view higher education as an organism with departments and colleges filling the niche of organs and organ systems. Living organ systems, of course, are also linked and are very resistant to change because of complex homeostatic mechanisms that are in place. For better or worse, all institutions of higher education I have observed are, likewise, very resistant to change because of homeostatic mechanisms they have developed, i.e., faculty senates, tenure procedures, curriculum committees, department and college autonomy, and the culture itself in higher education.

However, over the past two decades I have witnessed some changes to the higher education world. These changes have been brought about by a variety of accreditation, reaccredidation, assessment, and evaluation processes driven mainly by external sources. But more recently, internal forces are developing that are also bringing change. Having lived through two cycles of Southern Association of Colleges and Schools (SACS) reaffirmation for accreditation, first as Assistant Vice President for Academic Affairs and, more recently, as Dean of the College of Science, I have seen change occur.

I experienced, as Vice President for Academic Affairs, visits by three national accrediting bodies in the professional colleges, the Accreditation Board for Engineering and Technology (ABET), the National League for Nursing (NLN), and the American Association of Collegiate Schools of Business (AACSB). Each of these bodies has historically set rigid standards for accreditation, which, at times, I was convinced were designed to coerce administrators into providing higher salaries, additional faculty, and unlimited travel and operating budgets for the professional colleges. The Colleges of Liberal Arts and Science have never engaged in the process and the College of Education at UTEP withdrew from the accreditation process as did many other Schools and Colleges of Education throughout the country.

I have also been exposed to evaluation in my role as a faculty member and principal investigator or co-principal investigator on grant programs from the National Science Foundation (NSF) and the National Institutes of Health (NIH). The NSF has, in my opinion, been a major driving force in bringing change through assessment. I will discuss these changes later.

Finally, I have been able to view the changes that have occurred or sometimes not

occurred, at other institutions of higher education because of assessment and evaluation in my role as a site visitor and ad hoc consultant for a federal funding agency or consulting organization. Thus, I have, viewed assessment through the eyes of a faculty member, a chair, a dean, a vice president, and a consultant. It should come as no surprise that the views were, and will continue to be, different. Yet, in the world of higher education, we must somehow link these different views if we are to use assessment and evaluation as tools to bring about systemic change. That is, the criteria must be those that will move a department, a faculty, and an institution in the direction prescribed by the assessment and evaluation. Going back to my original analogy we must bring change to a loosely linked system of organs and organ systems that make up higher education.

The question to be addressed is, "What are the relationships between the evaluation criteria used by departments and institutions and those used by accrediting bodies, funding agencies, and our various professional learned societies?" We also need to ask whether these criteria can be used to bring about change in our approach to student learning, research, and service. The answer to the above two questions has too often been a negative one. Many faculty will state that there is no relationship between departmental and external criteria for evaluation and that the criteria used by SACS accrediting bodies are not being used to drive change in higher education systems. Some would argue that homeostasis prevails and the reward system is off target, but I am seeing change.

Coming back to SACS reaffirmation. In 1986, I served as co-chair of a committee on Undergraduate Educational Programs. A colleague, then chair of a department in Engineering, and I wrote the report and submitted it to a senior faculty member who was given the task of assembling the three volume SACS Reaccredidation Report. She received release time to manage this task; a modest committee worked through the process, the volumes were edited, the site committee appeared, we had a pleasant dinner at a local country club and nothing changed. All was well.

My second experience with SACS occurred eight years later. The previous report volumes were removed from the shelf, dusted off, and revisited. Low and behold, we had, in 1986, committed to a planning and evaluation process and made reference to outcomes assessment over the next 10-year period. Oh, really? The good news was that, under new leadership, by 1994 a Center for Institutional Evaluation, Research, and Planning (CIERP) had been established, a Director with the appropriate skills had been recruited, and serious thought was going into how a self-study would be conducted that involved faculty and professional staff across our campus, as opposed to the study that was mandated from the top in 1986. Over the next two years (1994-96) a self-study was conducted. The process actually began in fall 1993 when the CIERP, with support from the College of Business Administration, sponsored a series of electronic meetings for different groups to brainstorm the critical issues facing the university.

This information was shared with the Self-Study Advisory Committee that was appointed by the President in January 1994. This committee consisted of 58 members and met throughout the spring and summer of 1994 to examine the University Mission Statement, to discuss current legislative and community responsibilities, and to project trends in student enrollment. As a result of the committee's work, an expanded statement of purpose was drafted and disseminated across the campus. A series of eight campus forums followed in fall 1994 to inform faculty, staff, administrators, students, and community representatives about the self-study and to invite review of the draft statement. A 25-member Steering Committee was formed and worked with 13 subcommittees. The subcommittees completed their initial round of data

collection and analysis in spring 1995, and the Steering Committee met throughout the summer of that year to assess the criteria review process, to identify gaps and coverage, and to locate additional sources of information. By January 1996, a working draft of the self-study was widely disseminated across the campus. After ample feedback and modification, five copies of the draft SACS Self-Study were placed in the library, and the report was made available electronically for review and comment. Throughout January and February of 1996, several steering committee dialogues were established throughout the campus to discuss the Self-Study report. The Advisory Committee carefully studied all feedback and a final draft was prepared in February 1996. The final report was presented to the president in March 1996 and the SACS Visiting Committee spent three days of that same month on campus. Their efforts resulted in 22 recommendations and 17 suggestions. This, in my opinion, reflected two long years of work and self-reflection and indicated to me that my academic world was healthy. When one considers the 66 pages of Criteria for Accreditation in the 1992-94 edition of the SACS publication, only 22 recommendations and 17 suggestions must be viewed as a positive event.

How did the faculty respond to two years of self-study? Not well in many cases. Eyes rolled or glazed over at the mere mention of SACS and Assessment. One department in my college simply stated that they were accredited nationally. They taught the appropriate courses and gave grades. What other assessment criteria would I want? I got no response when I asked, "What evidence do you have that your students are successful once they obtained the baccalaureate degree?" As I viewed this second experience with SACS reaccredidation through the eyes of a dean, I felt that the exercise, although tedious at times, was worth the effort. We had attempted to engage a large number of faculty in the study and make them a part of it. I must say, however, that my feelings as a dean, were not shared by many of my faculty colleagues. In other words, the criteria used for SACS reaccredidation are not well incorporated into the minds of the faculty at the departmental level. The culture in my world is such that, although a logical argument can be mounted, we must look at outcomes assessment and success of our students and we must also think more about the learning process; unfortunately, the bottom line for many faculty is still their publication record and grants funded.

My perceptions of the previously mentioned accreditation visits and the criteria used by the professional colleges have been quite different from the faculty perceptions. As vice president, I worked with the dean of the College of Nursing and Health Sciences as we prepared for our NLN accreditation visit. The criteria were straightforward and we met them. If the criteria were not met or we were marginally meeting them, changes were made to comply with the NLN criteria for reaccredidation. The reaccredidation process in the College of Engineering was, likewise, rather straightforward. The criteria were clearly spelled out and any shortcomings perceived by the dean of the College of Engineering were addressed. In several cases, additional dollars were allocated to the College for equipment or other related expenses for student learning. My experience with AACSB accreditation was somewhat different, as our College of Business Administration had not been accredited. We moved considerable resources into the College to prepare for our initial accreditation visit. In our efforts to meet the AACSB criteria, new faculty were recruited, old faculty were in some cases moved aside, additional resources were allocated for library resources, and a reasonable, and in some people's minds unreasonable, amount of blood letting occurred. We did gain our initial accreditation and have since been reaccredited. As a result of the AACSB accreditation, the salary structure in the College of Business Administration today is

rather astonishing compared to what it was two decades ago. The issue of market-driven salaries is a point of contention between the College of Business and the other colleges. However there is no doubt that the criteria established by AACSB have brought change to the College of Business Administration.

I have also viewed assessment as a principal investigator or co-principal investigator on large NSF grants. One of these is a grant I shared with the dean of the College of Education (A Collaborative for Excellence in Teacher Preparation) and the other is a large university grant (Model Institutions for Excellence). In both cases, one of the objectives of the grant was to make major changes to the curriculum and especially to the way the faculty viewed student learning. Evaluation criteria on both of these grants include numbers of students impacted and, more importantly, through my eyes, whether or not the curriculum changes and the changes in the way course material was being delivered were making a difference in student learning. We did know whether or not more students were engaged in science, engineering, or mathematics, but in the case of the CETP Grant we needed to know whether or not more students were choosing to pursue careers in K-12 teaching of science and mathematics, whether or not they were better prepared teachers, and whether their students, in fact, were learning more due to the changes we made in the preservice preparation. When I say we, of course, I am referring to the faculty in the Colleges of Liberal Arts and Science, since they teach the majority of the hours required for preservice teachers. A major change has been made in the curricula used for preservice elementary preparation because of the grant. After considerable discussion between Education, Science, and Liberal Arts faculty and mentor teachers in the major school districts, 28 semester hours of math and science are required for certification. This is in contrast to the former degree plans that required only 15 hours of math and science. The courses utilized in the College of Science have also been redesigned to better meet the objectives of the preservice program. Thus, these grants have impacted not just the College of Education but the College of Liberal Arts and the College of Science, because the criteria used for evaluation were spread across the campus. I think I can safely say that at least a third of the faculty in the Colleges of Education, Liberal Arts, and Science have been impacted by the criteria used for evaluation of the programs, and certainly I can document the changes in our curriculum. I also can provide information that our students' attitudes toward science are changing. On the other hand, the SACS Self-Study, although viewed through my eyes as a positive, yet sometimes painful process, is still viewed by far too many faculty as a long, painful and perhaps over-designed process that they hope will sit on a shelf for the next 10 years. But again, through my eyes, as a dean, I spent far more time in four years discussing evaluation and outcomes and working with my peers, the deans of Liberal Arts and Engineering, than any deans I observed during the previous two decades on my campus. The story is far from over in the world of higher education, and my optimism is not shared by all on the campus. Of the 400 full-time tenured and tenure-track faculty, perhaps a third are cognizant of the criteria used by the accrediting bodies and of their concern with student outcomes assessment.

A final note. Other criteria used in Texas to evaluate the effectiveness of the university are set by the state legislature. Like most state governing bodies, they have set performance measures or criteria that are viewed as measures of success. The university is now impacted by output measures including numbers of undergraduate degrees, numbers of students needing remediation, percent of lower division courses taught by tenured or tenure-track faculty, percent of students completing courses, and percent of freshman who graduate within six years. The dollar amounts of externally funded research and

the pass rates for various licensure examines in Nursing, Education, and Engineering are also reported. These criteria may be on the minds of the average faculty member in the college involved, but are not factored into the general faculty annual evaluations. Thus again, I would have to conclude that these criteria are not completely congruent with those used at the departmental level. On the other hand, these criteria mandated through the legislative process obviously can bring a change to the university through the state funding formula.

What are the lessons learned? If assessment, evaluation, and accreditation are to be used as drivers for change in higher education, these processes must be truly shared and embraced by all members of the academy: students, faculty, and administrators. Everyone must be a full partner, but the linkages between the partners need not be extremely rigid.

When and if all the organs and organ systems are fully functioning in the assessment process, the appropriate changes will occur in the homeostatic mechanisms that control higher education.

An Assessment Model to Drive Undergraduate Educational Reform in the SMET Fields in a Large Public Multicampus University System

Manuel Gómez
Vice President for Research and Academic Affairs
Director of the Resource Center for Science and Engineering
University of Puerto Rico

Educational reform at the undergraduate level requires an institutional cultural transformation. Individual efforts of reform-oriented proactive faculty are necessary, but not sufficient. If a true reform is going to take place and is to be sustained, a systemic strategic plan that analyzes the whole system and identifies the strengths and weaknesses of the educational pipeline is needed. The plan also must identify the key pressure points that will catalyze the desired change and mechanisms that will nurture, protect, and motivate the agents of change and increase their capacity to influence the rest of the system.

The University of Puerto Rico (UPR) is a multicampus system with a 68,000 student body, eight 2- and 4-year colleges, and three graduate, Ph.D.-offering, campuses. A two pronged approach was found to be essential to guide, encourage, and nurture the reform across the system. A virtual organization, located in the office of the President of the University System, was established as a Reform Institute parallel to the regular academic management structure and designed to interact closely with top management and cadres of reform-oriented faculty.

The approach consisted of two major thrusts. The first was directed to the CEOs of the different campuses of the System, with the expressed objectives of (1) providing key systemic evaluation indicators that measure the effectiveness and efficiency of the undergraduate educational enterprise and (2) influencing major policy decisions that would institutionalize and accelerate educational reform. The second was directed to faculty, with the objective of nurturing the formation of a coherent cadre of reform-oriented professors who would experiment with new teaching/learning approaches, pilot test them, and then spearhead major systemic reform efforts.

This two-pronged approach was orchestrated by the academic management of the Resource Center for Science and Engineering (RCSE). Acting as a virtual organization, the Resource Center obtained external funds, mostly from the NSF through its Alliances for Minority Participation (AMP) project, to energize and catalyze the reform; conducted an extensive evaluation of the effectiveness and efficiency of the teaching/learning environment and activities; identified weaknesses and strengths; and promoted a strategic plan that would exploit key pressure points for initiating the reform. The RCSE then forged strategic alliances of reform-oriented faculty and nurtured pilot projects that experimented with new teaching/learning strategies grounded on the latest cognitive science and educational literature and on exemplary national projects. Simultaneously, the Center pioneered the development of metrics and benchmarks to measure the performance of the teaching/learning enterprise of the University System, measured these key systemic variables, and used them to persuade the CEO of the different campuses of the need for reform.

The result of this initiative, now in its second five-year phase, has been an institutional culture transformation of the

educational enterprise in key units of the system. CEOs of the academic units have been persuaded of the importance of institutional research to measure progress and locate points of ineffectiveness or inefficiency, guide strategic planning, and allocate resources on the basis of quality and output of department and colleges. Two of the campuses, well-conceived Faculty Development programs have been implemented to improve faculty teaching/learning skills and strategies. Institutional resources have been allocated to scale up successful pilot projects pioneered by the cadre of reform-minded faculty, and a program is being instituted for the professional development of the middle academic management of the institution to be effective supporters of the reform. The effort is known as the Academy for the Improvement of the Middle Management Support of the reform (AIMMS).

After five years of the reform effort, the number of science, mathematics, engineering, and technology (SMET) graduates per year has gone from 1,709 in 1991 to 2,674 in 1996--an increase of 56% in five years; this has been achieved without a significant increase in SMET enrollment, thus reflecting an increase in the effectiveness and efficiency of the educational process. These changes have been driven by a systematic measurement of graduation rates (GR), that have increased during the five-year period, and by the assessment of the effectiveness and efficiency of the gatekeeper courses in the SMET disciplines measured by an index of course efficiency (ICE), which has driven efforts to revise gatekeeper and bottleneck courses. The quality of graduates from the SMET programs has been measured through a proxy indicator or variable, the number of SMET Bachelor's recipients who pursue and complete a Ph.D. in these fields. One of the campuses has achieved what can be considered an outstanding result, even when benchmarked against national statistics; ten out of every hundred of its graduates from the B.S. programs go on to complete a Ph.D. in SMET at some of the top Ph.D. graduating institutions of the nation.

Key Systemic Metrics and Benchmarks as Vehicles to Drive Reform

When the AMP project started six years ago, the Institution had few systemic indicators, and these were not used in any systematic fashion by CEOs to guide their decision- and policymaking processes; thus, decision making was based mostly on intuition, guess work, and realpolitik. The AMP project developed a strategic plan to remedy this deficiency and achieve systemic reform in SMET programs. It started by studying the undergraduate SMET pipeline and finding its major ineffective and inefficient points; a series of courses were identified as the major source of the problem. Two categories of courses were identified: the gatekeeper courses (Pre-calculus and Calculus, Introductory Chemistry and Physics), where the greatest attrition of SMET students took place; and the bottleneck upper division or upper level courses that students had difficulty in approving and prevented SMET majors from graduating.

Multicampus assessment teams were assembled to evaluate the curriculum-content, teaching/learning strategies and methods, and classroom assessment of the gatekeeper courses. The teams then did an extensive evaluation of these courses and drew, by consensus, major recommendations for their improvement. Cadres of reform-oriented faculty were identified and pilot projects initiated to improve the teaching/learning process in these courses. Working groups were also formed to introduce cooperative learning methods in both gatekeeper and bottleneck courses, and a special program was developed for at-risk students—those students who, by the use of several indicators, had a high probability of dropping out of the SMET pipeline to develop their study/learning skills within the context

of a course (TADDEI, by its Spanish acronym).

Following the pipeline metaphor, a longitudinal cohort study was undertaken to identify the major obstacles to graduation and the nature of these impediments. Student records, not including student names to protect their privacy rights, of a scientifically selected random sample of SMET students from all campuses of the University of Puerto Rico, were studied and analyzed by campus and discipline. From this longitudinal cohort study, the following useful information was obtained: average time to graduate, year-to-year retention rates, percentage of cohort in good standing as they moved through the pipeline, *graduation rate*, percentage of SMET students who transferred to other disciplines, and average number of attempts needed to pass gatekeeper and bottleneck courses satisfactorily (A, B, or C). The *SMET graduation rate*—the percentage of entering SMET students who managed to graduate in seven or fewer years since admission (average time to graduate plus two years)—and the index of course efficiency (ICE)—the number of students in the cohort who took a specific course (including multiple attempts), divided by the number of students in a cohort who satisfactorily passed that course (thus, an ICE of 1 would mean that every student passed the course on the first attempt, and an index of 2 would mean that a student on the average would pass the course, after the second enrollment)—were the two most useful indicators for persuading the upper academic management, as well as faculty, of the need for reform. The ICE indicator, for example, convinced the Chancellors of two of the UPR campuses to institute Faculty Development programs and to establish special incentives to improve courses with high ICE numbers (high ICE indicates low efficiency).

To measure the quality of graduates, a proxy variable was designed that measured the number of students who completed a Ph.D. in SMET after graduating with a B.S. in one of these disciplines from the university. The data for this study were obtained from the National Academy of Sciences - National Research Council, Ph.D. study. Other indicators of quality were obtained by anecdotal means from interviews of B.S. graduates from UPR. A more systematic study of the performance, after graduation, of a random cohort of graduates is needed and will be included as part of the Institutional Research initiative that is being developed. Also included in the metrics was the total number of SMET degrees by campus, discipline, and gender, as required by NSF. These indicators--which had a specific goal of 2,600 SMET B.S. graduates for the first five years and of 4,000 for the second phase of AMP project--have been powerful tools for approving key policies and driving the reform.

Whenever possible, national benchmarks were identified to serve as indicators of progress in achieving the goals of the reform. For example, the flagship programs of Engineering (Mayagüez Campus) and Natural Sciences (Río Piedras Campus) have been benchmarked against graduation rates of the University of Illinois at the Urbana Campus. The Engineering School has already achieved its goal of 76% of the graduation rate of that Institution.

The rigorous assessment of the pilot reform projects has been an essential element to persuade CEOs and faculty to adopt strategies and methods pioneered by the pilot courses and to scale up these courses to department or collegewide level using institutional resources. Evidence of increased performance of pilot courses—using cooperative learning; TADDEI program; integration of laboratory and class, with emphasis on the development of concepts in physics; use of technology and innovative teaching strategies in the Pre-calculus/Calculus sequence; and a conceptually based hands-on course in Chemistry that also revised the class assessment tools to emphasize depth of understanding—are being used to drive the

reform and have helped set institutional strategic planning on a rational basis.

Systemic Outcomes and Conclusions

Based on an educational pipeline model, the SMET fields at the University were analyzed and a strategic plan was implemented to increase the effectiveness and efficiency of the educational enterprise and to transform the teaching/learning institutional culture. The reform was driven by a carefully designed assessment system that concentrated on the development of systemic metrics and outcomes. Carefully articulated goals were enunciated, and key benchmarks identified to measure progress. The focus of the reform initiative was located at the Resource Center for Science and Engineering, an organization within the university system that supersedes the traditional departmental, college, and campus academic management structures. It has been successful in gaining the endorsement of the upper academic management of the university of the merits of its strategic plan and systemic assessment of the educational enterprise. The Center has pioneered a model institutional research program that brings together strategic planning, institutional assessment, targeted pilot projects to spearhead reform, and the establishment of policy to allocate infrastructure and human resources to achieve agreed upon goals.

As a result of this initiative and through careful design of systemic metrics and benchmarks, key weaknesses in the educational pipeline were identified in the form of SMET gatekeeper and bottleneck courses. Pilot projects to reform the system were pioneered.

Evidence was provided to persuade top decision makers of the importance of institutional research as the appropriate mechanism to establish policies following a rational approach that results in the improvement of the effectiveness and efficiency of the educational system. Cadres of reform-oriented educators were nurtured and supported in the development of pilot projects to reform undergraduate education. Evaluation of these pilot projects has persuaded CEOs of at least two campuses to scale up the reform. They have also implemented Faculty Development Programs for the improvement of faculty teaching/learning skills and methods. At the system level, an Academy for the Improvement of the Middle Management to Support the Educational Reform is being developed.

The reform effort in SMET has spilled over to the area of Teacher Preparation (TP) programs. By joining efforts, the AMP project and the Puerto Rico Statewide Systemic Initiative (PR-SSI) have harnessed a strategic alliance of the schools of Education and Natural Sciences to reform the SMET-TP programs. The assessment methodology of the AMP project described in this paper has been adapted to pioneer a similar educational reform of TP programs with the active participation of reformed teachers and schools of the PR-SSI.

The Resource Center for Science and Engineering has helped to set a Science and Technology Policy for the government of Puerto Rico. The Policy calls for strengthening human resources development capacity of the University to meet the challenge of the knowledge economy. Following this Policy will ensure that the goal of awarding 4,071 B.S. degrees, by the year 2001 from Puerto Rican universities will be achieved, a net increase of 138% from the base year of 1991 (1,709 graduates).

This goal should be reached as a result of an increase in the effectiveness of the teaching/learning enterprise and a transformation of the institutional culture at UPR. The University now has as its mission to develop and strengthen its capacities as a research/teaching institution with equal weight given to both components.

Technology-Assisted Learning in Higher Education: It Requires New Thinking About Assessment

Sheri D. Sheppard
Associate Professor, Mechanical Engineering—Design Division
Co-Director of the Stanford Learning Lab
Stanford University

The purpose of this paper is to outline some of the issues, challenges, and questions facing universities and colleges as they consider the use of technology in the support of teaching and learning. The issues are difficult, in that they require balanced consideration of questions such as:

> What types of technologies improve learning?
> Which faculty members will adopt and experiment with technologies? How will this impact the reward system?
> Which technologies can we afford? What are the hidden costs?

Most of these questions remain unanswered.

We begin by giving examples of the types of learning technologies that universities are exploring and adopting. We proceed to enumerate some of the reasons for these adoptions and then discuss the various groups who should be asking probing questions about the effectiveness of the technologies. Some of the characteristics of an evaluation plan that would address these various questions are then proposed. We conclude with several emerging models for such an evaluation plan.

Examples of Adoption

Over the last ten years there has been growing interest in universities and colleges across the U.S. in the use of technology for supporting learning in higher education.

- The Anderson Center for Innovation in Undergraduate Education at Rensselaer Polytechnic Institute is creating computer-based learning environments that enable students to get multiple views of science concepts and provide training on the use of the World Wide Web in instruction. (*http://ciue.rpi.edu/about.html*)
- The University of Illinois is exploring the role of asynchronous learning for undergraduates to improve cost effectiveness and efficiency of education delivery. (*http://w3.scale.uiuc.edu/scale/*)
- Stanford University has created an extensive menu of graduate level engineering courses available synchronously to enable students around the world to work on a master's degree. (*http://scpd.stanford.edu/news.html#scpd*)
- The University of Oklahoma has created multimedia-based learning experiences to supplement traditional course experiences (e.g., strength of materials, dynamics) and also to create nontraditional experiences (e.g., a trip to mars). (*http://eml.ou.edu/*)
- Both the University of Illinois and Stanford University offer subsets of campus-based courses online. (*http://www.online.uillinois.edu/*) (*http://stanford-online.stanford.edu/*)
- The NSF Synthesis Coalition maintains a Web-based library of engineering curricular materials from which faculty from across the country can "borrow" and to which faculty can submit. (*http://www.needs.org/*)

More and more university courses across the nation are supported with Web sites, newsgroups, and discussion forums. One need only look at a copy of the magazine *Syllabus* to

get a sense of the explosion in the use of learning technologies. In this paper, we define *learning technology* as any learning tool that uses computers or advanced communication systems.

What Is Motivating This Interest?

There are a number of reasons that universities are considering and adopting the use of learning technologies. Some of them are related to beliefs that learning technologies:

- are inherently "good"
- are needed to remain competitive as an institution
- make the delivery of education more cost effective
- open up possibilities of reaching new/different student groups
- offer students more control over when and where they interact with "knowledge"
- offer more and new opportunities for student-student and student-faculty interaction
- offer new opportunities for cross-university interaction by both students and faculty
- offer students richer, more diverse learning resources and alternate points of view
- offer opportunities for documenting, cataloguing, and re-use of curriculum materials and student work.

All of these factors, except the first three, are focused on improving student learning and/or the faculty work environment.

Who Needs Evaluation Results?

Evaluation of the contributions and limitations of technology in supporting and improving learning are sorely needed. The need exists for all participants. For example, *university administrators* need evaluation results in order to be able to make informed financial and policy decisions. On the financial front, university administrators need to decide how many resources to direct toward creating, maintaining, and updating a technology infrastructure in their institution. On the policy front, the university needs, for example, to decide how to view faculty participation in creating technology-based course resources when it comes time for tenure and promotion, or to decide which groups of students they want to better serve with technology tools.

In addition, *faculty* need evaluation results in order to make more informed decisions about how they might use technology in their teaching. *Students* need evaluation results to be able to make more informed decisions about what types of learning environments work best with their learning styles.

Of course, no single evaluation study could address all of the questions posed by the various groups listed above, in terms of the level of detail, point of view, and relevant data. In spite of these differences, all of these groups are in need of constructive feedback that can only result from thoughtfully posed, implemented, and assessed curricular experiments.

Difficult Problem

Assessment to evaluate technology effectiveness in promoting learning is difficult. It is problematic because participants and stakeholders range from students and faculty, to school administrators, to university service groups. In addition, it is hard because the types of questions that these groups ask on the relative value of technology in improving learning are quite diverse. For example, a school administrator may ask, "What are the relative advantages of providing Internet connections to all dorm rooms vs. providing better staffing in the dorm computer clusters?" An art professor may ask, "Will students utilize the course slide collection more extensively over the Internet than when it was on reserve in the library, even though the visual quality is inferior? What are the copyright issues related to some of the images in the collection?" And an engineering professor may ask, "Are there more effective means of teaching difficult concepts? Are there

advantages to students being able to turn in their problem sets digitally? How will I provide them with feedback?"

While it is important to address and answer these individual questions, it is equally important for a university community to come to an understanding of the role of technology in supporting the work of students and faculty. Assessment and evaluation need to be looked at in new and extended ways.

- *Assessment must be undertaken at a more minute level in learning activities in order to discover what elements of an academic experience contributed most to learning effectiveness.* This means, among other things, that faculty and students be more reflective about rationale for their actions. So of course, both faculty and student buy-in to assessment is critical. It is not enough to simply ask students at the end of a term, "What is your overall rating of this course/instructor?" In many institutions, including Stanford, student surveys are generally the only measure used to evaluate course quality. The results of these surveys are not timely, often being returned to the instructor several months after the term is over, and are so abbreviated as to give little direction for course improvement/changes.

- *Assessment must be seen as a partnership activity.* University administrators, faculty, and students all need to embark on the sort of assessment described, above believing that its primary objective is the improvement of learning, and that they are all stakeholders in discovering and validating what elements work and, similarly, what elements do not. Faculty and students alike need this partnership to support their taking thoughtful risks in their teaching and learning. If the assessment is at all perceived to be judgmental, the partnership will break down.

- *Assessment must be seen as a long-term commitment by all of the partners and as a way of doing business.* School administrators, faculty, and students alike need to adopt the attitude that improvement is continuous (i.e., what I learn this year goes into changes for next year), and that some significant effects on student learning may be measured over years, rather than over a single quarter.

- *Assessment results must be synthesized, publicized, and disseminated.* If individual faculty continue to look only at results from their own curricular experiments (with and without technology), there is no opportunity to view the cumulative effect of an education on students. The university needs to take responsibility for synthesizing individual results. In addition, the metalevel findings and results should be disseminated to faculty within and beyond the institution.

Emerging Assessment Models

There are a number of technology-related projects and programs in colleges and universities across the country that are developing Evaluation/Assessment models that fulfill all or some of the requirements outlined in the previous section. A few of these models will be discussed here. This discussion is not intended to be comprehensive, but rather illustrative. We start by offering several examples of projects that are *undertaking assessment at a more minute level in learning activities in order to discover what elements of an academic experience contributed most to its effectiveness.* The papers of Reamon and Sheppard (1996, 1997) and Regan and Sheppard (1996) are focused on developing an understanding of the role of simulation software/courseware and physical models in affecting students' understanding of mechanical systems. In the former, the "mechanical system" was 4-bar linkages, and

in the latter it was the drivetrain of a bicycle. Both projects utilized an assessment technique called "video interaction analysis" that involves videotaping student learning activities so that a diverse group of researchers could then repeatedly revisit the activity to fully examine, from multiple perspectives, what students were doing (Jordan & Henderson, 1992/in press). In this way, a detailed understanding of the roles that the simulation software/courseware played in the student learning can emerge. Both projects used small sample sizes (less than 15 student volunteers). The more recent work of D. Reamon (NSF grant DUE-9653114) is looking at the role of simulation software in larger samples of student learning. This work focuses more specifically on the role of "interactivity" in promoting conceptual understanding and information retention. Assessment instruments that are being used in the two-week long experiment include pre- and posttests, surveys, and video interaction analysis. The sample size is 105 junior and senior mechanical engineering students taking a required course for their major.

A final example of projects exploring at a minute level the interactions of students and technology comes not from the engineering domain, but from the humanities domain. In fall 1997 the Stanford Learning Laboratory (SLL) piloted a course called "Introduction to Humanities: The Word and the World" (*http://sll1.stanford.edu/wordworld/index.html*). As mandated by the Faculty Senate, the course focused on introducing freshmen to the methods of inquiry that researchers and scholars in the humanities use to study five significant texts. The quarter-long course was supported by a Web-based backbone, and conventional lectures and discussion sections. The course's Web environment provided content, supporting and enrichment materials on all five texts; posed and collected short-answer questions prior to each lecture; and provided both lecture and section forums. In parallel with the creation of the Web backbone during summer 1997, the Stanford Learning Laboratory funded the work of an Assessment Team, which became known as the A-Team. The A-Team of six core individuals and five advisors worked in partnership with the teaching and technology teams to define teaching and learning issues, goals, and questions that motivated the investment in Web technologies for the course and to develop an appropriate assessment plan (i.e., one that would address questions being posed by the various stakeholders—SLL, the teaching team, and the Faculty Senate). The plan utilized a variety of assessment tools (e.g., surveys, interviews, video taping, Web statistics) that allowed for direct evaluation of utilization and effectiveness of particular elements of the Web backbone and allowed for directly addressing questions posed by the stakeholders. At the same time, the tools afforded a broader window into the utilization of technology more generally by the 90 Stanford undergraduates in the course.

It is harder to find examples of learning-technology projects that are explicitly promoting assessment to *be a partnership activity between the University, faculty and students; a long-term commitment by all of the partners—a way of doing business;* and/or *an undertaking resulting in synthesis, publishing and disseminating of results.* Fortunately we can report that a few are emerging. For example, the Sloan Project at the University of Illinois (*http://w3.scale.uiuc.edu/scale/*) is collecting, analyzing, and publishing assessment data from both faculty and students for many of the technology assisted courses that it is creating.

Another example is at the Anderson Center for Innovation in Undergraduate Education at Rensselaer Polytechnic Institute, which is actively researching how interactive learning has improved students' education or cut additional costs. Rensselaer has recently been awarded a grant from the Atlantic Philanthropic Service Company that will enable Rensselaer Polytechnic Institute to follow two groups of RPI students throughout the second half of their undergraduate education and into the early years of their

careers or graduate studies. One group will have taken interactive studio courses and the other primarily traditional lecture-based courses. A detailed cost analysis of interactive vs. traditional education is also being carried out as part of this study by Rensselaer economists.

A third example is the Stanford Learning Laboratory (SLL), which was established in 1997 by Stanford University President Gerhard Casper and the Commission on Technology in Teaching and Learning (*http://learninglab.stanford.edu/*). Its mission is to enhance the personal learning experience of all Stanford students and to create a model for the judicious deployment of pedagogically informed technology for learning and knowledge management. In all projects, the SLL conducts a comprehensive benchmark-assessment using qualitative and quantitative methods to evaluate project effectiveness, utility, impact, and deployment barriers.

We offer a final example that does not fall within the traditional definition of a university—the NSF-sponsored Synthesis Coalition. Synthesis represents a partnership of eight universities exploring the use of technology to improve learning in engineering schools. Synthesis has set as its goal not only the creation and dissemination of engineering courseware through a distributed Web-based database (*http://www.needs.org/*), but also the assessment of the courseware by developing and promoting quality standards and recognition for outstanding courseware development (e.g., the John Wiley Premier Award).

All of these projects face major challenges in gaining and sustaining faculty and student buy-in and interest, in balancing needs for minute-level assessment data with policy level information, and in finding the appropriate voices for getting information and findings back to the various constituents (e.g., university administrators, faculty, students) and to the broader university community across the country.

Closing Remarks

Technology-based learning tools, whether they are Web-based courses, electronic newsgroups, or simulation-based courseware, all hold the potential for improving student learning. It is imperative, however, that we embark on the exploration and adoption of any of these tools in a thoughtful manner. We, as faculty and administrators, should be asking hard questions about the relative merit of the tools. Many of these questions will be unanswerable until we undertake well-posed curricular experiments and pilot studies. A large component of any of these curricular experiments should be an assessment plan that addresses questions about technology and learning of concern to faculty, students, and university administrators alike.

References

Jordan, G., & Henderson, A. (in press). Interaction analysis: Foundations and practice (A report for Xerox Palo Alto Research Center and Institute for Research on Learning, 1992). *The Journal of the Learning Sciences.*

Reamon, D., & Sheppard, S. D. (1996). Analytic problem solving methodology. In *IEEE Proceedings from 1996 Frontiers in Education Conference* (vol.1, pp. 484-488). Nov. 6-9, Salt Lake City, Utah.

Reamon, D., & Sheppard, S. D. (1997). The role of simulation software in an ideal learning environment. In *Proceedings of the ASME Design Theory and Methodology Conference*, Sept. 14-17.

Regan, M., & Sheppard, S. D. (1996). Interactive multimedia courseware and hands-on learning experience: An assessment study. *ASEE Journal of Engineering Education,* 85(2), 123-130.

Appendix B
Participant List

Judy Ackerman
Montgomery College
51 Mannakee Street
Rockville, MD 20850
Ph. 301.279.5027
Fax 301.279.5028
jackerma@mc.cc.md.us

Madeline Adamczeski
Chemistry Department
American University
4400 Massachusetts Avenue, NW
Washington, DC 20016-8014
Ph. 202.885.1761
Fax 202.885.1752
madamcz@american.edu

Clifford Adelman
U.S. Department of Education
Office of Educational Research & Improvement
555 New Jersey Avenue, NW, 617A
Washington, DC 20208
Ph. 202.219.2251
Fax 202.219.2030
cadelman@inet.ed.gov

Baine Alexander
LEAD Center
University of Wisconsin–Madison
1402 University Ave, Room 435
Madison, WI 53703
Ph. 608.265.5921
Fax 608.265.5923
baine@engr.wisc.edu

Charles Allen
Mathematics Department
Drury College
900 N. Benton
Springfield, MO 65802
Ph. 417.873.7210
Fax 417.873.7432
callen@lib.drury.edu

Marwan Amarin
Biology/Natural Science
Richard J. Daley College
7500 S. Pulaski Road
Chicago, IL 60652
Ph. 773.838.7730
Fax 773.838.7524

Virginia Anderson
Biology Department
Towson University
Towson, MD 21252
Ph. 410.830.3041
Fax 410.830.2405
vanderson@towson.edu

David Andrews
Department of Biology
California State University–Fresno
2555 E. San Ramon
Fresno, CA 93740
Ph. 209.278.2412
Fax 209.278.3963
david_andrews@csufresno.edu

Rose Asera
The Charles A. Dana Center
University of Texas at Austin
2901 N. IH 35, Suite 3.200
Austin, TX 78722
Ph. 512.475.9715
Fax 512.471.6193
asera@mail.utexas.edu

Kevin Aylesworth
American Physical Society
Education & Outreach Programs
One Physics Ellipse
College Park, MD 20740
Ph. 301.209.3245
Fax 301.209.0865
ayleswor@aps.org

Christina Bailey
Chemistry & Biochemistry California Polytechnic State University
San Luis Obispo, CA 93407
Ph. 805.756.2443
Fax 805.756.5500
cbailey@calpoly.edu

Richard Baldwin
Technologies/Health Care
Quinsingamond Community College
607 West Boylston
Worchester, MA 01606
Ph. 508.854.4364
Fax 508.852.6943
richardb@qcc.mass.edu

Collin Ballance
Mathematics & Computer Information Systems Departments
Nashville State Technical Institute
120 White Bridge Road
Nashville, TN 37209
Ph. 615.353.3278
Fax 615.353.3428
ballance_c@nsti.tec.tn.us

Robert Barak
State Board of Regents
East 12th and Grand Avenue
Old Historical Building
Des Moines, IA 50319
Ph. 512.281.3934
Fax 512.281.6420
rbarak@iastate.edu

Robert Barkley
National Education Association
National Center for Innovation
1201 16th Street, NW
Washington, DC 20036
Ph. 202.822.7370
Fax 202.822.7482
bbarkley@nea.org

Lehman Barnes
Department of Natural Science
University of North Florida
4567 St. Johns Bluff Road
Jacksonville, FL 32224
Ph. 904.620.1074
Fax 904.620.1025
lbarnes@unf.edu

Marianne Barnes
Department of Curriculum & Instruction
University of North Florida
4567 St. Johns Bluff Road
Jacksonville, FL 32224
Ph. 904.620.2578
Fax 904.620.1025
mbarnes@unf.edu

James Barufaldi
Science Education Center
University of Texas at Austin
1912 Speedway, SZB 340
Austin, TX 78712
Ph. 512.471.7354
Fax 512.471.9244
james@mail.utexas.edu

David Bauman
Capital Area Institute for Mathematics & Science
55 Miller Street
PO Box 489
Summerdale, PA 17093
Ph. 717.732.8427
Fax 717.732.8414
dbauman@caiu.k12.pa.us

Nancy Baxter Hastings
Mathematics & Computer Science
Dickinson College
PO Box 1773
Carlisle, PA 17013-2896
Ph. 717.245.1626
Fax 717.245.1690
baxter@dickinson.edu

Jean Beard
Biological Sciences
San Jose State University
One Washington Square
San Jose, CA 95192
Ph. 408.924.4870
Fax 408.924.4840
beard@biomail.sjsu.edu

Ray Beiersdorfer
Geology Department
Youngstown State University
One University Plaza
Youngstown, OH 44555
Ph. 330.742.1753
Fax 330.742.1754
ray@cc.ysu.edu

Spencer Benson
Microbiology Department
University of Maryland–College Park
College Park, MD 20742
Ph. 301.405.5478
Fax 301.314.9489
sb77@umail.umd.edu

Andrew Bernat
Model Institutions for Excellence
University of Texas at El Paso
El Paso, TX 79968
Ph. 915.747.8888
Fax 915.747.5243
abernat@utep.edu

Janet Boese
American Chemical Society
Education Division
1155 Sixteenth Street, NW
Washington, DC 20036
Ph. 202.872.6164
Fax 202.833.7732
j_boese@acs.org

Jack Bookman
Mathematics Department
Duke University
PO Box 90320
Durham, NC 27708-0320
Ph. 919.660.2831
Fax 919.660.2821
bookman@math.duke.edu

Don Bord
Department of Natural Sciences
University of Michigan-Dearborn
4901 Evergreen Road
Dearborn, MI 48128
Ph. 313.593.5483
Fax 313.593.4937
dbord@sb-f1.umd.umich.edu

Dianne Bowcock
LEAD Center
University of Wisconsin–Madison
1402 University, Room 413
Madison, WI 53706
Ph. 608.265.5924
Fax 608.265.5923
bowcock@engr.wisc.edu

Myles Boylan
National Science Foundation
Division of Undergraduate Education
4201 Wilson Boulevard, Suite 835
Arlington, VA 22230
Ph. 703.306.1665
Fax 703.306.0445
mbiylan@nsf.gov

Thomas Brady
4720 N. Mesa, #28
El Paso, TX 79912
Ph. 915.533.6603

Stacey Lowery Bretz
Department of Natural Sciences
University of Michigan–Dearborn
4901 Evergreen Road
Dearborn, MI 48128
Ph. 313.593.5157
Fax 313.593.4937
slbretz@umich.edu

Jack Bristol
Department of Biological Sciences
University of Texas–El Paso
El Paso, TX 79968
Ph. 915.747.5844
Fax 915.747.5808
jbristol@utep.edu

Ted Britton
National Center for Improving Science Education
2000 L Street, NW, Suite 616
Washington, DC 20036
Ph. 202.467.0652
Fax 202.467.0659
britton@ncise.org

Aaron Brower
Social Work
University of Wisconsin–Madison
1350 University Avenue
Madison, WI 53706
Ph. 608.263.3838
Fax 608.263.3836
ambrower@facstaff.wisc.edu

Bonnie Brunkhorst
Department of Geological Science
California State University–San Bernardino
5500 University Parkway
San Bernardino, CA 92407
Ph. 909.880.5612
Fax 909.780.3640
bbrunkho@wiley.csusb.edu

Herbert Brunkhorst
Department of Biology
California State University–San Bernardino
5500 University Parkway
San Bernardino, CA 92407
Ph. 909.880.5613
Fax 909.880.5988
hkbrunkh@wiley.csusb.edu

Merle Bruno
Natural Science
Hampshire College
West Street
Amherst, MA 01002
Ph. 413.582.5414
Fax 413.582.5448
mbruno@hampshire.edu

Paul Bucci
Academy for Education
Higher Education Management Services
1255 23rd Street
Washington, DC 20037
Ph. 202.884.8158
Fax 202.884.8466
pbucci@aed.org

David Buchthal
Buchtel College of Arts & Sciences
University of Akron
Akron, OH 44325
Ph. 330.972.7880
Fax 330.972.7222
dbuchthal@uakron.edu

Ann Burgess
Biology, Core Curriculum University of Wisconsin–Madison
361 Noland Hall
250 North Mills Street
Madison, WI 53706-1794
Ph. 608.263.1594
aburgess@facstaff.wisc.edu

Daniel Burke
National Science Foundation
Office of Assistant Director, Education & Human Resources
4201 Wilson Boulevard
Arlington, VA 22230
Ph. 703.306.1602
Fax 703.306.0399
dburke@nsf.gov

Gail Burrill
National Council of Teachers of Mathematics
1906 Association Drive
Reston, VA 20191-1593
Ph. 703.620.9840
Fax 414.425.6442
gburrill@macc.wisc.edu

Judith Burry-Stock
Educational Research
University of Alabama
Box 870231
327 Graves Building
Tuscaloosa, AL 35487
Ph. 205.348.1187
Fax 205.348.0683
jburry@bamged.va.edu

Bruce Callen
Physics Department
Drury College
900 N. Benton
Springfield, MO 65802
Ph. 417.873.7473
Fax 417.873.7432
bcallen@lib.drury.edu

Barbara Cambridge
American Association for Higher Education
Assessment Forum
One Dupont Circle, Suite 360
Washington, DC 20036
Ph. 202.293.6440
Fax 202.293.0073
bcambridge@aahe.org

Charles Cannon
Science & Mathematics
Columbia College Chicago
600 S. Michigan Avenue
Chicago, IL 60605
Ph. 312.344.7396
Fax 312.341.0542
ccannon@popmail.colum.edu

Joseph Cannon
Chemical Engineering
Howard University
2300 6th Street NW
Washington, DC 20059
Ph. 202.806.6626
Fax 202.806.4635
jcannon@scs.howard.edu

Wayne Carley
National Association of Biology Teachers
11250 Roger Bacon Drive, #19
Reston, VA 20190
Ph. 703.471.1134
Fax 703.435.5582
wcarley@aol.com

Nancy Carnal
Biology Department
San Francisco State University
1600 Holloway Avenue
San Francisco, CA 94132
Ph. 415.338.1853
Fax 415.338.6816
ncarnal@sfsu.edu

Rita Caso
TX Engineering Experiment Station
TX AMP/Foundation Coalition
015 WERC (Bizzell Street)
College Station, TX 77843-3405
Ph. 409.862.4375
Fax 409.862.1267
mrc1586@unix.tamu.edu

Lillian Cassel
National Science Foundation
Directorate for Education & Human Resources
4201 Wilson Boulevard
Arlington, VA 22230
Ph. 703.306.1669
Fax 703.306.0445
lcassel@nsf.gov

Wanda Chambers
National Institute on Student Achievement, Curriculum, & Assessment
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, DC 20208-5573
Ph. 202.219.2035
Fax 202.219.2133
wanda_chambers@ed.gov

Amy Chang
American Society for Microbiology
Board of Education & Training
1325 Massachusetts Avenue, NW
Washington, DC 20005
Ph. 202.942.9264
Fax 202.942.9329
achang@asmusa.org

Norman Chonacky
The Evergreen State College
Mail Stop L 3220
Olympia, WA 98502
Ph. 315.866.6000 x5028
chonacky@evergreen.edu

Daryl Chubin
National Science Foundation
Directorate for Education & Human Resources
4201 Wilson Boulevard, #855S
Arlington, VA 22230
Ph. 703.306.1650 x5801
Fax 703.306.0434
dchubin@nsf.gov

Naomi Chudowsky
National Academy of Sciences
Board on Testing & Assessment
2101 Constitution Avenue, NW
Washington, DC 20418
Ph. 202.334.1455
Fax 202.334.3584
nchudows@nas.edu

Julia Clark
U.S. House of Representatives
1641 Longworth Building
Washington, DC 20515
Ph. 202.225.4531
Fax 202.225.5662
julia.clark@mail.house.gov

Robert Clark
Department of Physics
Texas A&M University
College Station, TX 77843
Ph. 409.845.3332
Fax 409.845.2590
rbc@amu.edu

William Clune
National Institute for Science Education
1025 West Johnson
Madison, WI 53706
Ph. 608.263.4348
clune@macc.wisc.edu

Elaine Jane Cole
Center for Science Education
Portland State University
Box 751
Portland, OR 97210
Ph. 503.725.8763
Fax 503.725.3884
colee@psu4.pdx.edu

Brian Coppola
Department of Chemistry
University of Michigan
930 North University Avenue
Ann Arbor, MI 48109-1055
Ph. 734.764.7329
Fax 734.647.4865

Donald Cotten
Center for Science & Mathematics Education
The University of Southern Mississippi
Box 5087
Hattiesburg, MS 39406
Ph. 601.266.4739
Fax 601.266.4739

Frank Creegan
Chemistry Department
Washington College
300 Washington Avenue
Chestertown, MD 21620
Ph. 410.778.7725
Fax 410.778.7276
frank.creegan@washcoll.edu

Francena Cummings
Eisenhower Consortium for Mathematics & Science
345 S. Magnolia Drive, E-22
Tallahassee, FL 32301
Ph. 850.671.6033
Fax 850.671.6010
fdc3530@garnet.acns.fsu.edu

Sandra Cynar
Computer Engineering & Computer Science Department
CSULB
1250 Bellflower Boulevard
Long Beach, CA 90840
Ph. 562.985.1512
Fax 562.985.7823
cynar@engr.csulb.edu

Judy Da Walt
Madison Area Technical College
3550 Anderson Street
Madison, WI 53704
Ph. 608.246.6680
Fax 608.246.6783
jmd9052@madison.tec.wi.ue

Charlene D'Avanzo
School of Natural Science
Hampshire College
Amherst, MA 01002
Ph. 413.582.5569
Fax 413.582.5448
cdavanzo@hampshire.edu

Kerry Davidson
Louisiana Board of Regents
Louisiana Systemic Initiatives Program
150 3rd Street., Suite 129
Baton Rouge, LA 70801
Ph. 504.342.4253
Fax 504.342.6926
davidson@regents.state.la.us

Norma Davila
Resource Center for Science & Engineering
University of Puerto Rico
PO Box 23334
San Juan, PR 00931-3334
Ph. 787.765.5170
Fax 787.756.7717
n_davila@upr1.upr.clu.edu

Donald Deeds
Biology Department
Drury College
900 N. Benton
Springfield, MO 65802
Ph. 417. 873.7398
Fax 417.873.4732
ddeeds@lib.drury.edu

Robert DeHaan
Department of Cell Biology
Emory University Health Science Center
575 Rollins Way
Atlanta, GA 30322
Ph. 404.727.3050
Fax 404.727.3051
rld@cellbio.emory.edu

Connie Della-Piana
Model Institutions for Excellence
University of Texas at El Paso
500 West University Avenue
Geology Room 124
El Paso, TX 79968
Ph. 915.747.8888
Fax 915.747.5243
connie@utep.edu

Gabriel Della-Piana
El Paso Collaborative for Academic Excellence
University of Texas at El Paso
413 Education Bldg.
El Paso, TX 79968
Ph. 915.747.5155
Fax 915.747.5144
gabriel@utep.edu

Katherine Denniston
Biology Department
Towson University
8000 York Raod
Towson, MD 21252
Ph. 410.830.3128
Fax 410.830.2405
kdenniston@towson.edu

Nancy Devino
National Research Council
Center for Science, Mathematics & Engineering Education
2101 Constitution Avenue, NW
Washington, DC 20418
Ph. 202.334.1462
Fax 202.334.3159
ndevino@nas.edu

Terry Devitt
National Institute for Science Education
Office of News & Public Affairs
500 Lincoln Drive
Madison, WI 53706
Ph. 608.262.8282
Fax 608.262.2331
trdevitt@facstaff.wisc.edu

LaDonna Dickerson
National Center for Improving Science Education
2000 L. Street, Suite 616
Washington, DC 20036
Ph. 202.467.0652
Fax 202.467.0659
dickerso@ncise.org

Kathy DiRanna
WestEd
California Schools Implementation Network
University of California–Irvine
Irvine, CA 92697-4680
Ph. 714.824.7809
Fax 714.824.7621
kathy_diranna@cams.edu

Peter Dorhout
Chemistry Department
Colorado State University
Fort Collins, CO 80523
Ph. 970.491.0624
Fax 970.491.1801
pkd@lamar.colostate.edu

Lloyd Douglas
National Science Foundation
Division of Mathematics Sciences
4201 Wilson Boulevard, Room 1025
Arlington, VA 22230
Ph. 703.306.1874
Fax 703.306.0555
ldouglas@nsf.gov

Don Duggan-Haas
Teacher Enhancement
Michigan State University
301 D. Erickson Hall
East Lansing, MI 48824
Ph. 517.355.1725
Fax 517.432.5092
haasdona@pilot.msu.edu

Hubert Dyasi
School of Education
City College of New York
Convent Avenue and 136th Street
New York, NY 10031
Ph. 212.650.8436
Fax 212.650.6292
xdyasi@aol.com

Rebecca Dyasi
Education Department
City College of New York
Convent Avenue and 136th Street
New York, NY 10031
Ph. 212.650.6263
Fax 212.650.6292
redcc@cunyvm.cuny

Julie Ealy
Chemistry/Science Education
Columbia University
509 W. 121 Street
Bancroft 606
New York, NY. 10027
Ph. 212.678.3575
jbe10@columbia.edu

Janice Earle
National Science Foundation
Elementary, Secondary & Informal Education
4201 Wilson Boulevard, Room 885
Arlington, DC 22230
Ph. 703.306.1614
Fax 703.306.0412
jearle@nsf.gov

Diane Ebert-May
Science & Mathematics Learning Center & Department of Biological Sciences
Northern Arizona University
Box 5697
Flagstaff, AZ 86011
Ph. 520.523.9125
Fax 520.523.7953
diane.ebert-may@nau.edu

Betty Eidemiller
American Society for Microbiology
Board of Education & Training
1325 Massachusetts Avenue, NW
Washington, DC 20005
Ph. 202.942.9299
Fax 202.942.9329
beidemiller@asmusa.org

Arthur Ellis
University of Wisconsin–Madison
College Level One
7351 Chemistry Building
Madison, WI 53706
Ph. 608.262.0421
Fax 608.262.6143
ellis@fozzie.chem.wisc.edu

Nicholas Eror
University of Pittsburgh
Associate Dean's Office
323 Benedum Hall
Pittsburgh, PA 15261
Ph. 412.624.9761
Fax 412.624.1108
eror+@pitt.edu

Eugenia Etkina
Graduate School of Education
Rutgers, The State University of New Jersey
10 Seminary Place
New Brunswick, NJ 08901-1183
Ph. 732.932.7496 x339
Fax 732.932.7552
etkina@rci.rutgers.edu

Joyce Evans
National Science Foundation
Teacher Enhancement
4201 Wilson Boulevard, Room 885
Arlington, VA 22230
Ph. 703.306.1620
Fax 703.306.0412
jevans@nsf.gov

Davene Eyres
Science & Math Division
North Seattle Community College
9600 College Way North
Seattle, WA 98103
Ph. 206.528.4515
Fax 206.527.3748
deyres@sccd.ctc.edu

Roya Farhoosh
Chemistry Department
Columbia University
22 B Anton Road
Storrs, CT 06268
Ph. 860.429.9001
rfarhoosh@worldnet.att.net

Lorraine Fleming
Civil Engineering
Howard University
2300 6th Street, NW
Washington, DC 20059
Ph. 202.806.6570
Fax 202.806.5271
fleming@scs.howard.edu

Claudia Fley
Arlee High School
Science Department
874 Detweiler Road
Arlee, MT 59821
Ph. 406.726.3211
Fax 406.726.3216
arl4332@montana.net.internet

Julie Foertsch
LEAD Center
University of Wisconsin–Madison
1402 University Avenue, Room 435
Madison, WI 53703
Ph. 608.265.6368
Fax 608.265.5923
foertsch@engr.wisc.edu

Norman Fortenberry
National Science Foundation
Division of Undergraduate Education
4201 Wilson Boulevard, Suite 835
Arlington, VA 22230
Ph. 703.306.1670
Fax 703.306.0445
nfortenb@nsf.gov

Scott Franklin
Department of Physics & Astronomy
Dickinson College
HUB/College and Louther Streets
Carlisle, PA 17013
Ph. 717.245.1797
Fax 717.245.1642
franklis@dickinson

Joseph Frattaroli
The Teachers Academy for Mathematics & Science in Chicago
3424 S. State Street
Chicago, IL 60616
Ph. 312.949.2422
Fax 312.808.0103
jfrattaroli@tams.iit.edu

David Fromson
School of Natural Science & Mathematics
California State University, Fullerton
PO Box 6850
Fullerton, CA 92834-6850
Ph. 714.278.2638
Fax 714.278.5390
dfromson@fullerton.edu

Helene Gabelnick
Physical Science Department Harold Washington College
30 E. Lake Street
Chicago, IL 60601
Ph. 312.553.5787
Fax 312.553.5964

Tom Gadsden
Eisenhower National Clearinghouse
Ohio State University
1929 Kenny Road
Columbus, OH 43210
Ph. 614.292.3330
Fax 614.292.2066
tgadsden@enc.org

Susan Ganter
Science Service
Science Education Programs
1719 N. Street, N.W.
Washington, DC 20036
Ph. 202.785.2255
Fax 202.785.1243
sganter@sciserv.org

David Garin
Chemistry Department
University of Missouri-St. Louis
8001 Natural Br. Road
St. Louis, MO 63121
Ph. 314.516.5349
Fax 314.516.5342
garin@umsl.edu

Todd Gary
NSF Local Systemic Change Initiative in Science Education
Tennessee State University
330 10th Avenue, N.
Box 141 Suite J
Nashville, TN 37203
Ph. 615.963.7219
Fax 615.963.7214
tgary@picard.tnstate.edu

Mike Gehner
Biology Department
Xavier University
3800 Victory Parkway
Cincinnati, OH 45207
Ph. 513.745.2055
Fax 513.745.1954
gehner@xavier.xu.edu

Esther Gibbs
Chemistry Department
Goucher College
1021 Dulaney Valley
Baltimore, MD 21204
Ph. 410.337.6338
Fax 410.337.6408
egibbs@goucher.edu

Katy Ginger
University Corporation for Atmosphere Research
Program for the Advancement of Geoscience Education
3450 Mitchell Lane
Boulder, CO 80301
Ph. 303.497.8341
Fax 303.497.8336
ginger@ucar.edu

Ellen Goldstein
School of Education
City College of New York
138th Street Convent Avenue
New York, NY 10031
Ph. 212.650.6700
Fax 212.650.6221
gold3100@con2.com

Bernard Goldstein
California State University
Board of Trustees
111 Park Avenue
San Carlos, CA 94070
Ph. 650.591.2576
Fax 650.591.2527

James Golen
Chemistry & Biochemistry
UMASS Dartmouth
N. Dartmouth, MA 02747
Ph. 508.999.8245
Fax 508.999.9167
jgolen@umassd.edu

Manuel Gómez
Resource Center for Science & Engineering
University of Puerto Rico
P.O. Box 23334
San Juan, PR 00931
Ph. 787.764.8369
Fax 787.756.7717
m_gomez@uprl.upr.clu.edu

David Gosser
Chemistry Department
The City College of CUNY
138th St. and Convent Avenue
New York, NY 10031
Ph. 212.650.8375
Fax 212.650.8339
gosser@sci.ccny.cuny.edu

Ramona Gunter
LEAD Center
University of Wisconsin–Madison
1402 University Avenue
Madison, WI 53706
Ph. 608.262.9514
Fax 608.265.5923
rgunter@engr.wisc.edu

Larry Gursky
Ronan Middle School
Science
Ronan, MT 59864
Ph. 406.676.3390
Fax 406.676.3393
larryg@cyberport.net

Joshua Gutwill
Department of Chemistry
University of California–Berkeley
ModularChem Consortium
Berkeley, CA 94720
Ph. 510.643.5610
Fax 510.643.1471
gutwill@socrates.berkeley.edu

Kamel Haddad
Mathematics Department
California State University–Bakersfield
9001 Stockdale Hwy.
Bakersfield, CA 93311
Ph. 805.664.2150
Fax 805.664.2039
kdaddad@ultrix6.csubak.edu

Alfred Hall
AEL-Arlington
Eisenhower Regional Consortium for Math & Science
1700 N. Moore Street, Suite 1275
Arlington, VA 22209-1903
Ph. 703.558.2246
Fax 703.276.0266
halla@ael.org

Loren Hall
Ivy Tech State College
Planning & Education
One West 26th Street
Indianapolis, IN 46208
Ph. 317.921.4987
Fax 317.921.4629
lhall@ivy.tec.in.us

Kimberly Hambrick
Appalachia Educational Laboratory, Inc.
Eisenhower Regional Mathematics & Science Consortium
1031 Quarrier Street
Charleston, WV 25301
Ph. 304.347.1888
hambrick@aol.org

Josephine Hamer
Mathematics & Computer Science Department
Western Connecticut State University
181 White Street
Danbury, CT 06810
Ph. 203.837.9347
Fax 203.837.8339
hamer@wcsu.ctstateu.edu

Bo Hammer
American Institute of Physics Education
One Physics Ellipse
College Park, MD 20740
Ph. 301.209.3013
Fax 301.209.0839
bhammer@aip.org

Vivian Hampton
College of Engineering
North Carolina A&T State University
1601 East Market Street
Greensboro, NC 27411
Ph. 910.334.7447
Fax 910.334.7540
vivian@ncat.edu

David Hanson
Chemistry Department
State University of New York
Stony Brook, NY 11794-3400
Ph. 516.632.7917
Fax 516.632.7960
david.hansen@sunysb.edu

David Hata
Microelectronics Technology Portland Community College
18624 NW Walker Road
Beaverton, OR 97006
Ph. 503.533.2929
Fax 503.533.2948
dhata@pcc.edu

Jacqueline Haynes
Intelligent Automation, Inc.
2 Research Place
Rockville, MD 20850
Ph. 301.590.3155
Fax 301.590.9414
jhaynes@i-a-i.com

Teresa Hein
Physics Department
American University
4400 Massachusetts Avenue, NW
Washington, DC 20016-8058
Ph. 202.885.2766
Fax 202.885.2723
thein@american .edu

Warren Hein
American Association of Physics Teachers
One Physics Ellipse
College Park, MD 20740
Ph. 301.209.3323
Fax 301.209.0845
whein@aapt.org

Lars Helgeson
Teaching & Learning
University of North Dakota
PO Box 7189
Grand Forks, ND 58202
Ph. 701.777.3144
Fax 701.777.4393
lhelgeso@badlands.nodak.edu

Michael Henle
Mathematics Department
Oberlin College
Oberlin, OH 44074
Ph. 216.775.8380
Fax 216.775.6638
michael.henle@oberlin.edu

James Highsmith
Office of the Chancellor
California State University
Academic Senate
400 Golden Shore, Suite 132
Long Beach, CA 90802
Ph. 562.985.2613
Fax 562.985.2618

Susan Hixson
National Science Foundation
Division of Undergraduate Education
4201 Wilson Boulevard
Arlington, VA 22230
Ph. 703.306.1667
Fax 703.306.0445
shixson@nsf.gov

Daniel Householder
National Science Foundation
Division of Elementary, Secondary & Informal Education
4201 Wilson Boulevard, Room 885
Arlington, VA 22230
Ph. 703.306.1620
Fax 703.306.0412
dhouseho@nsf.gov

Barbara Howard
Programs in Clinical Laboratory Science
Catholic University of America
Room 11, McCort-Ward Biology Building
Washington, DC 20064
Ph. 202.319.5270
Fax 202.319.5721
howardb@cua.edu

Karen Hubbard
Biology Department
The City College of New York
138th Street at Convent Avenue
New York, NY 10031
Ph. 212.650.8566
Fax 212.650.8585
khubbard@scisun.sci.ccny.cuny.edu

Mary Ann Huntley
National Center for Improving Science Education
2000 L. Street, Suite 616
Washington, DC 20036
Ph. 202.467.0652
Fax 202.467.0659

Ben Hutchinson
Chemistry Department
Pepperdine University
24255 Pacific Coast Highway
Malibu, CA 90263
Ph. 310.456.4331
Fax 310.456.4785
bhutchin@pepperdine.edu

Ann Igoe
MATEC
4225 N. 21st Street, #4
Phoenix, AZ 85016
Ph. 602.589.2416
Fax 602.589.2532
aigoe@aol.com

M. Frank Ireton
American Geophysical Union
2000 Florida Avenue, NW
Washington, DC 20009
Ph. 202.462.6900
Fax 202.328.0566
fireton@kosmos.agu.org

Marilyn Irving
Curriculum & Instruction
School of Education
Howard University
2441 4th Street, NW
Washington, DC 20059
Ph. 202.806.7339
Fax 202.806.5297
mirving@fac.howard.edu

Tony Jacob
National Institute for Science Education
Chemistry Learning Center
1101 University Avenue
Madison, WI 53706
Ph. 608.263.5647
atjacob@facstaff.wisc.edu

Kamil Jbeily
Texas Regional Collaboratives for Excellence in Science Teaching The University of Texas at Austin
Science Education Center
1912 Speedway SZB 340
Austin, TX 78712
Ph. 512.471.9460
Fax 512.471.9244
kjbeily2mail.utexas.edu

Jean Johnson
National Science Foundation
Science Resources Studies, Room 965
4201 Wilson Boulevard
Arlington, VA 22230
Ph. 703.306.1780
Fax 703.306.0510

Elmima Johnson
National Science Foundation
Directorate for Education & Human Resources
4201 Wilson Boulevard
Arlington, VA 22230
Ph. 703.306.1650
Fax 703.306.0456
ejohnson@nsf.gov

Donald Jones
National Science Foundation
Elementary, Secondary & Informal Education
4201 Wilson Boulevard
Arlington, VA 22230
Ph. 703.306.1620
Fax 703.306.0412
djones@nsf.gov

Trace Jordan
Morse Academic Plan
New York University
269 Mercer Street, Room 804
New York, NY 10003
Ph. 212.998.078
Fax 212.995-4055
trace.jordan@nyu.edu

Alex Kajstura
Science, Engineering & Technology
Palm Beach Community College
4200 Congress Avenue
Lake Worth, FL 33461-4705
Ph. 561.439.8131
Fax 561.439.8255
kajstura_a@popmail.firn.edu

Conrad Katzenmeyer
National Science Foundation
REC Room 855
4201 Wilson Boulevard
Arlington, VA 22230
Ph. 703.306.1653
Fax 703.306.0434
ckatzenm@nsf.gov

Kylie Keshav
Bioscience & Biotechnology
Drexel University
32nd and Chestnut Streets
Philadelphia, PA 19104
Ph. 215.895.1972
Fax 215.895.1273
keshavkf@dunxl.ocs.drexel.edu

Timothy Killeen
Space Physics Research
University of Michigan
2455 Hayward
Ann Arbor, MI 48843
Ph. 734.936.2745
Fax 734.763.0437
tkilleen@umich.edu

Athanassios Kodogeorgiou
Natural Science
R. Daley College
7500 S. Pulaski Road
Chicago, IL 60652
Ph. 773.838.7717
Fax 773.838.7524
tomk2@pipeline.com

Jane Korey
Mathematics Department
Dartmouth College
Bradley Hall 6188
Hanover, NH 03755
Ph. 603.646.1048
Fax 603.646.1312
jane.lcorey@dartmouth.edu

Adrienne Kozlowski
Connecticut Academy for Education in Math, Science & Technology
211 South Main Street
Middletown, CT 06457
Ph. 860.346.1177
Fax 860.346.2157
kozlowskia@ccsua.ctstate.edu

Paul Kuerbis
Education Department
Colorado College
14 E. Cache La Poudre Street
Colorado Springs, CO 80903
Ph. 719.389.6726
Fax 719.389.6473
pkuerbis@cc.colorado.edu

Jay Labov
National Research Council
Center for Science, Mathematics, & Engineering Education
HA 450
2101 Constitution Avenue, NW
Washington, DC 20418
Ph. 202.334.1458
Fax 202.334.3159
jlabov@nas.edu

Carole Lacampagne
U.S. Department of Education
National Institute Post Secondary Education
555 New Jersey Avenue, NW
Washington, DC 20208
Ph. 202.219.2064
Fax 202.219.2030

Phil LaFontaine
WestEd
California Schools Implementation Network
CSU Sacramento
6000 J Street - FF/1
Sacramento, CA 95819-6120
Ph. 916.278.5927
Fax 916.278.4770

Joan LaFrance
Mekinak Consulting
4710 Aurora N., #202
Seattle, WA 98103
Ph. 206.547.6904
Fax 206.684.5809
jlafrance@aol.com

Diana Lambdin
Curriculum & Instruction–School of Education
Indiana University
Wright Education Bldg., Room 3266
201 North Rose Avenue
Bloomington, IN 47405
Ph. 812.856.8149
Fax 812.856.8116
lambdin@indiana.edu

Neal Lane
National Science Foundation
4201 Wilson Boulevard
Arlington, VA 22230
Ph. 703.306.1000
Fax 703.306.1869
nlane@nsf.gov

Helen Lang
Department of Philosophy
Trinity College
Hartford, CT 06106
Ph. 860.297.2419
Fax 860.297.5358
helen@lang@trincoll.edu

James Lankford
Chemistry Department
St. Andrews Presbyterian College
1700 Dogwood Mile
Lauriniburg, NC 28352
Ph. 910.277.5321
Fax 910.277.5020
jil@tartan.sapc.edu

Sandra Laursen
Department of Chemistry
Kalamazoo College
1200 Academy Street
Kalamazoo, MI 49006
Ph. 616.337.7020
Fax 616.337.7251
laursen@kzoo.edu

Priscilla Laws
Department of Physics & Astronomy
Dickinson College
HUB/College & Louther Street
Carlisle, PA 17013
Ph. 717.245.1242
Fax 717.245.1642
lawsp@dickinson.edu

Roger Lederer
College of Natural Sciences
California State University–Chico
Chico, CA 95929
Ph. 530.898.6121
Fax 530.898.4363
rlederer@oavax.csuchico.edu

Dennis Lehman
Physical Science
Harold Washington College
30 E. Lake Street
Chicago, IL 60601
Ph. 312.553.5794
dlehman@ccc.edu

Zafra Lerman
Institute for Science Education & Science Communication
Columbia College–Chicago
600 S. Michigan Avenue
Chicago, IL 60605-1996
Ph. 312.334.7180
Fax 312.663.5172
zafral@aol.com

Herb Levitan
National Science Foundation
Division of Undergraduate Education
4201 Wilson Boulevard
Arlington, VA 22230
Ph. 703.306.1666
Fax 703.306.0445
hlevitan@nsf.gov

Stephen Lewis
National Institute for Science Education
Professional Audiences
2000 L. Street, Suite 616
Washington, DC 20036
Ph. 202.467.0652
Fax 202.467.0659
lewis@ncise.org

Eileen L. Lewis
Cañada College
4200 Farm Hill Boulevard
Redwood City, CA 94061
Ph. 415.306.3255
Fax 510.841.3511
eileen@socrates.berkeley.edu

James Lightbourne
National Science Foundation
Division of Undergraduate Education
4201 Wilson Boulevard
Arlington, VA 22230
Ph. 703.306.1665
Fax 703.306.0448
jlight@nsf.gov

Robert Linck
Chemistry Department
Smith College
Northampton, MA 01063
Ph. 413.585.3836
Fax 413.585.3786
rlinck@sophin.smith.edu

Alberta Lipson
Massachusetts Institute of Technology
CDSUE
Room 20B-140
77 Massachusetts Avenue
Cambridge, MA 02139
Ph. 617.253.8604
Fax 617.253.0055
lipson@mit.edu

Kate Loftus
National Institute for Science Education
College Level One
798 Educational Sciences Building
1025 W. Johnson St.
Madison, WI 53706
Ph. 608.263.5681
Fax 608.262.7428
skloftus@facstaff.wisc.edu

Melanie Loo
Biological Sciences
CSU Sacramento
6000 J Street
Sacramento, CA 95819
Ph. 916.278.6573
mwloo@csus.edu

Susan Loucks-Horsley
NISE/National Research Council
Professional Development
4732 N. Oracle Road, Suite 217
Tucson, AZ 85705
Ph. 520.888.2838
Fax 520.888.2621
sloucks@wested.org

Kathy Luker
College of Engineering Student Affairs Office
University of Wisconsin–Madison
1415 Engineering Drive
Madison, WI 53706
Ph. 608.265.3761
Fax 608.262.6400
luker@engr.wisc.edu

Pam Magasich
National Council for Accreditation of Teacher Education
2010 Massachusetts Avenue, NW
Washington, DC 20036
Ph. 202.466.7496
Fax 202.466.6620
pam@ncate.org

Ted Maguder
Natural Science
St. Petersburg Junior College
2465 Drew Street
Clearwater, FL 33765
magudert@mail.spjc.cc.fl.us

Laura Markham
Chemistry Department
Michigan State University
119 Chemistry Building
East Lansing, MI 48823
Ph. 616.355.9715 x366
markham@pilot.msu.edu

Cora Marrett
University of Massachusetts–Amherst
Room 362
Whitmore Administration Building
Amherst, MA 01003
Ph. 413.545.2554
cmarrett@provost.umass.edu

Jill Marshall
Physics Department
Utah State University
UMC 4415, USU
Logan, UT 84322-4415
Ph. 435.797.2883
Fax 435.797.2492
marshall@cc.usu.edu

Sarah Mason
LEAD Center
University of Wisconsin–Madison
1402 University Avenue
Madison, WI 53706
Ph. 608.265.6370
Fax 608.265.5923
smason@facstaff.wisc.edu

Bob Mathieu
Astronomy Department
University of Wisconsin–Madison
475 N. Charter Street
Madison, WI 53706
Ph. 608.262.5679
Fax 608.263.0361
mathieu@astro.wisc.edu

Eric Mazur
Harvard University
29 Oxford Street
Cambridge, MA 02138
Ph. 617.495.8729
Fax 617.495.9837
mazur@phics.harvard.edu

Patricia McAllister
Educational Testing Service
1800 N. Street, Suite 900
Washington, DC 20036
Ph. 202.659.8042
Fax 202.659.8075
pmcallister@ets.org

Raymond McGhee
SRI International
Higher Education Policy
1611 N. Kent St.
Arlington, VA 22209
Ph. 703.247.8573
Fax 703.247.8493
mcghee@wdc.sri.com

Curtis McKnight
Mathematics Department
University of Oklahoma
601 Elm Avenue
Norman, OK 73019
Ph. 405.325.2728
Fax 405.325.7484
cmcknight@ou.edu

Flora McMartin
Synthesis Coalition
UC–Berkeley
3112 Etcheverry Hall
Berkeley, CA 94720
Ph. 510.643.2928
Fax 510.643.1822
mcmartin@synthesis.org

Joy McMillan
Agriscience, Apprenticeship, Technical & Industrial Division
Madison Area Technical College
3550 Anderson Street
Madison, WI 53704
Ph. 608.246.6001
Fax 608.246.6995
jmcmillan@madison.tec.wi.us

Cathy Middlecamp
Department of Chemistry University of Wisconsin–Madison
1100 University Avenue
Madison, WI 53706
Ph. 608.263.5647
Fax 608.262.0381
chmiddle@facstaff.wisc.edu

Chand Midha
Mathematics/Sciences
University of Akron
229 Ayer Hall
Akron, OH 44325
Ph. 330.972.7128
Fax 330.374.8630
cmidha@uakron.edu

Susan Millar
LEAD Center
University of Wisconsin–Madison
1402 University Avenue
Madison, WI 53706
Ph. 608.265.5943
Fax 608.265.5923
smillar@engr.wisc.edu

Terrance Millar
University of Wisconsin–Madison
500 Lincoln Drive
Madison, WI 53706
Ph. 608.263.1600
Fax 608.262.6400
millar@math.wisc.edu

Nancy Minear
California Alliance for Minority Participation
University of California–Irvine
600 Administration
Irvine, CA 92697
Ph. 714.824.2669
Fax 714.824.3048
naminear@uci.edu

Joel Mintzes
Department of Biological Sciences
University of North Carolina
601 S. College Road
Wilmington, DE 28403
Ph. 910.962.3437
Fax 910.962.4066
mintzes@uncwil.edu

Sue Mitchell
Computer Information Systems, Business Division
Calhoun Community College
PO Box 2216
Decatur, AL 35609
Ph. 205.306.2655
Fax 205.306.2506
sue@calhoun.cc.al.us

Loretta Molitor
Physics Department
Towson University
8000 York Road
Towson, MD 21252
Ph. 410.830.2532
Fax 410.830.3959
lmolitor@towson.edu

Jean Moon
Exxon Education Foundation
Mathematics Program
321 Reedy Meadow Road
Groton, MA 01450-1407
Ph. 978.448.0626
Fax 978.448.0626
mbb321@ultranet.com

Marlene Moore
College of Arts & Science
University of Portland
5000 N. Willamette Boulevard
Portland, OR 97203
Ph. 503.283.7221
Fax 503.283.7399
moorem@up.edu

Trish Morse
National Science Foundation
Elementary, Secondary & Informal Education
4201 Wilson Boulevard
Arlington, VA 22230
Ph. 703.306.1614
Fax 703.306.0412
mpmorse@nsf.gov

Gregory Moses
Engineering Physics
College of Engineering
University of Wisconsin–Madison
1415 Engineering Drive
Madison, WI 53706
Ph. 608.263.1600
Fax 608.262.6400
moses@engr.wisc.edu

William Mowczko
National Institutes of Health
Office of Science Education
6100 Executive Boulevard
Rockville, MD 20852
Ph. 301.402.5224
Fax 301.402.3034
billm@amb.niddk.nih.gov

Susan Mundry
National Institute for Science Education
91 Montvale Avenue
Stoneham, MA 02180
Ph. 781.279.8215
Fax 781.279.8220
smundry@wested.org

Paul Musial
Mathematics Department
Daley College
7500 S. Pulaski Road
Chicago, IL 60652
Ph. 773.838.7634
Fax 773.838.7524
pmusial@uic.edu

Terri Nally
American Chemical Society
Education Division
1155 Sixteenth Street, NW
Washington, DC 20036
Ph. 202.872.4587
Fax 202.833.7732
t_nally@acs.org

Roger Nanes
Physics Department
California State University–Fullerton
Fullerton, CA 92834
Ph. 714.278.2188
Fax 714.278.5810
rnanes@fullerton.edu

Paul Neill
Physics Department
University of Nevada–Reno
Reno, NV 89557
Ph. 702.784.1307
Fax 702.784.1398
paul@physics.unr.edu

Mary Neuman
Annenberg Institute for School Reform
Brown University
One Davol Square
Providence, RI 02903
Ph. 401.863.7970
Fax 401.863.1290
mary_neuman@brown.edu

Anthony Nicastro
Physics Department
West Chester University of Pennsylvania
West Chester, PA 19383
Ph. 610.436.2540
Fax 610.436.3013
anicastro@wcupa.edu

Kathleen O'Sullivan
Secondary Education
San Francisco State University
1600 Holloway Avenue
San Francisco, CA 94132
Ph. 415.338.1599
Fax 415.338.0914
kaosul@sfsu.edu

Charlotte Otto
Department of Natural Sciences
University of Michigan–Dearborn
4901 Evergreen Road
Dearborn, MI 48128
Ph. 313.593.5277
Fax 313.593.4937
cotto@umich.edu

Lynette Padmore
Biology Department
Florida A&M University
1540-G South Adams Street
Tallahassee, FL 32307
Ph. 850.561.2467
Fax 850.561.2684
lpadmore@ns1.famu.edu

Frank Palocsay
Chemistry MSC 7701
James Madison University
Harrisonburg, VA 22807
Ph. 540.568.8065
Fax 540.568.3581
palocsfa@jmu.edu

Carlo Parravano
Merck Institute for Science Education
PO Box 2000, RY60-215, 126 East Lincoln Avenue
Rahway, NJ 07065
Ph. 732.594.3443
Fax 732.594.3977
carlo_parravano@merck.com

R. B. Partridge
Astronomy Department
Haverford College
Haverford, PA 19041
Ph. 610.896.1144
Fax 610.896.4904
bpartrid@haverford.edu

Celeste Pea
National Science Foundation
Educational System Reform
4201 Wilson Boulevard, Suite 875
Arlington, VA 22230
Ph. 703.306.1682
Fax 703.306.0456
cpea@nsf.gov

G. Earl Peace
Department of Chemistry University of Wisconsin–Madison
1101 University Avenue
Madison, WI 53706
Ph. 608.262.8647
Fax 608.265.8094
gpeacejr@facstaff.wisc.edu

Judith Pelchat
Annenberg Institute for School Reform
Brown University
One Davol Square
Providence, RI 02903
Ph. 401.863.7562
Fax 401.863.1290
judith_pelchat@brown.edu

Barbara Pellegrini
846 W. Hart
Beloit, WI 53511
Ph. 847.467.2489
Fax 847.491.8999
bpellegr@inwave.com

Debby Penberthy
LEAD Center
University of Wisconsin–Madison
1402 University Avenue, Room 417
Madison, WI 53706
Ph. 608.265.5927
Fax 608.265.5923
penberth@engr.wisc.edu

Sarah Pfatteicher
LEAD Center
University of Wisconsin–Madison
1402 University Avenue
Madison, WI 53706
Ph. 608.265.5925
Fax 608.265.5923
spfatt@engr.wisc.edu

Stephanie Pfirman
Environmental Science
Barnard College
3009 Broadway
New York, NY 10027
Ph. 212.854.5120
Fax 212.854.5760
spfirman@barnard.columbia.edu

Andrew Porter
National Institute for Science Education
University of Wisconsin–Madison
1025 West Johnson Street
Madison, WI 53706
Ph. 608.263.4200
Fax 608.263.6448
acporter@macc.wisc.edu

Dave Porter
Department of Behavior Sciences & Leadership
U. S. Air Force Academy
2354 Fairchild Drive, Suite 6671
USAFA, CO 80840
Ph. 719.333.2514
Fax 719.333.3338
dave-porter@usa.net

J. Preston Prather
Center of Excellence in Science & Mathematics Education
University of Tennessee–Martin
145 Gooch Hall
Martin, TN 38238-5029
Ph. 901.587.7166
Fax 901.587.7206
jpprather@utm.edu

Nikki Privacky
Biology Department
Palm Beach Community College
4200 Congress Avenue
Lake Worth, FL 33461
Ph. 561.434.5120
Fax 561.434.5009

Senta Raizen
National Center for Improving Science Education
2000 L. Street, Suite 616
Washington, DC 20036
Ph. 202.467.0652
Fax 202.467.0659
raizen@ncise.org

Ann Redelfs
National Partnership in Advanced Computational Infrastructure
Eternal Relations
9500 Gilman Drive
La Jolla, CA 92093
Ph. 619.534.5032
Fax 619.534.5113
redelfs@sdsc.edu

Kimberly Reed
American Association of Physics Teachers
One Physics Ellipse
College Park, MD 20740
Ph. 301.209.3344
Fax 301.209.0845
kreed@aapt.org

Sharon Roberts
Department of Biology
California State University–Bakersfield
9001 Stockdale Highway
Bakersfield, CA 93309
Ph. 805.664.2220
Fax 805.664.2132
skroberts@csubak.edu

Wayne Roberts
Macalester College
1600 Grand Avenue
St. Paul, MN 55105
Ph. 612.696.6160
Fax 612.696.6075
robertsw@macalester.edu

Doris Roman
The Gem Consortium
Academic Programs
Building 410
1609 E. Helen Street
Tucson, AZ 85721
Ph. 520.626.5193
Fax 520.626.3277
droman.ll@nd.edu

Nancy Romance
Science Education
College of Education
Florida Atlantic University
777 Glades Road
Boca Raton, FL 33431
Ph. 361.297.3577
Fax 561.297.3794
romance@fau.edu

Nedah Rose
John Wiley & Sons, Inc.
605 Third Avenue
New York, NY 10158-0012
Ph. 212.850.6345
Fax 212.850.6591
nrose@wiley.com

Susan Ross
Mathematics Department
The University of Southern Mississippi
Box 5045
Hattiesburg, MS 39406
Ph. 601.266.4289
Fax 601.266.5818
sross@whale.st.usm.edu

Vicki Roth
Learning Assistance Services
University of Rochester
107 Lattimore Hall
Rochester, NY 14627
Ph. 716.275.9049
Fax 716.273.1116
vrth@uhura.cc.rochester.edu

George Rublein
Mathematics Department
College of William & Mary
Williamsburg, VA 23187-8795
Ph. 757.221.2028
Fax 757.221.2988
gtrabl@math.wm.edu

Joel Russell
Chemistry Department
Oakland University
Rochester, MI 48309
Ph. 248.370.2086
Fax 248.370.2321
russell@oakland.edu

Gerhard Salinger
National Science Foundation
Elementary, Secondary & Informal Science
4201 Wilson Boulevard
Arlington, VA 22230
Ph. 703.306.1620
Fax 703.306.0412
gsalinge@nsf.gov

Mathew Santirocco
College of Arts & Science
New York University
100 Washington Square East, Room 910
New York, NY 10003-6688
Ph. 212.998.8100
Fax 212.995.4141
cyberdean@nyu.edu

Steve Schneider
WestEd
4140 Jefferson Avenue
Woodside, CA 94052
Ph. 650.949.8470
Fax 650.949.8482
sschnei@wested.org

Patricia Schroeder
Science Health Care, Mathematics
Johnson County Community College
12345 College Boulevard
Overland Park, KS 66210
Ph. 913.469.8500 x3133
Fax 913.469.2518
pschroed@johnco.cc.ks.us

Suku Sengupta
Civil & Environmental Engineering
University of Massachusetts–Dartmouth
285 Old Westport Road
North Dartmouth, MA 02747
Ph. 508.999.8470
Fax 508.999.8964
ssengupta@umassd.edu

Elaine Seymour
Bureau of Socialogical Research
University of Colorado–Boulder
Campus Box 580
Boulder, CO 80309
Ph. 303.492.0084
Fax 303.492.2154
seymour@spot.colorado.edu

Lynn Shelby
Geosciences Department
Murray State University
PO Box 9
Murray, KY 42071-0009
Ph. 502.762.6761
Fax 502.762.4417
shelby@bach.mursuky.edu

Sheri Sheppard
Mechanical Engineering
Stanford University
16 Peter Courts Circle
Stanford, CA 94305
Ph. 650.725.1590
Fax 650.723.3521
sheppard@cdr.stanford.edu

Eleanor Siebert
Physical Sciences & Mathematics
Mount St. Mary's College
12001 Chalon Road
Los Angeles, CA 90049
Ph. 310.954.4101
Fax 310.954.4379
esiebert@msmc.la.edu

Mary Simmons
NSF Local Systemic Change Initiative in Science Education
Tennessee State University
330 10th Avenue North
Box 141 Suite J
Nashville, TN 37203
Ph. 615.963.7230
Fax 615.963.7214
bnya@picard.tnstate.edu

Patricia Simpson
Biology/Science Education
St. Cloud State University
720 Fourth Avenue South
St. Cloud, MN 56301
Ph. 320.255.3012
Fax 320.255.4166
psimpson@stcloud.edu

Charles Singler
Geology Department
Youngstown State University
One University Plaza
College of Arts & Sciences
Youngstown, OH 44555
Ph. 330.742.3611
Fax 330.742.1754
c.singler@ysu.edu

Christopher Sirola
South Carolina Center for Advanced Technological Education
Science
Highway 76
PO Box 587
Pendleton, SC 29670
Ph. 864.646.8361 x2325
Fax 864.646.8256
csirola@tricounty.tec.sc.us

Harold Skelton
Physics Department
West Chester University of Pennsylvania
West Chester, PA 19383
Ph. 610.436.3010
Fax 610.436.3013
hskelton@wcupa.edu

Doris Sligh
Office of Elementary and Secondary Education
Compensatory Education Programs
Team 1
U.S. Department of Education
11401 Symphony Wood Lane
Silver Spring, MD 20901
Ph. 202.260.0999
Fax 202.260.7764
doris_sligh@ed.gov

Sharron Smith
Chemistry & Physics
Hood College
401 Rosemont Avenue
Frederick, MD 21701
Ph. 301.696.3675
Fax 301.696.3367
ssmith@nimue.hood.edu

Barbara Spector
Secondary Education
University of South Florida
15836 Sanctuary Drive
Tampa Palms, FL 33647-1075
Ph. 813.971.1856
Fax 813.975.1015
spector@typhoon.coedu.usf.edu

Brock Spencer
Department of Chemistry
Beloit College
700 College Street
Beloit, WI 53511
Ph. 608.363.2249
Fax 608.363.2718
spencer@beloit.edu

Diane Spresser
National Science Foundation
Elementary, Secondary & Informal Education
4201 Wilson Boulevard
Arlington, VA 22230
Ph. 703.306.1613
Fax 703.306.0412
dspresse@nsf.gov

Leonard Springer
Wisconsin Center for Education Research
University of Wisconsin–Madison
1025 W. Johnson Street
Madison, WI 53706
Ph. 608.263.4254
Fax 608.262.7428
lspringe@facstaff.wsic.edu

Rita Starnes
WestEd
California Schools Implementation Network
34449 Fawn Lane
Squaw Valley, CA 93675
Ph. 209.332.2123
Fax 209.332.2324
rita_starnes@cams.edu

Donna Sterling
GSE
George Mason University
4400 University Drive
Fairfax, VA 22030
Ph. 703.993.2043
Fax 703.993.2013
dsterlin@gmu.edu

Joseph Stewart
National Science Foundation
Elementary, Secondary & Informal Education
4201 Wilson Boulevard, Room 885
Arlington, VA 22230
Ph. 703.306.1620
Fax 703.306.0412
jstewart@nsf.gov

James Stith
American Institute of Physics
One Physics Ellipse
College Park, MD 20740
Ph. 301.209.3126
Fax 301.209.0841
jstith@aip.org

Thomas Stoebe
Materials Science & Engineering
University of Washington
Box 532120
Seattle, WA 98195
Ph. 206.543.7090
Fax 206.543.3100
stoebe@u.washington.edu

Sarah Stoll
Chemistry Department
Oberlin College
130 West Lorain Street
Oberlin, OH 44074
Ph. 216.775.8741
Fax 216.775.6682
fstoll@oberlin.edu

Andrei Straumanis
Chemistry Department
Washington College
300 Washington Avenue
Chestertown, MD 21620
Ph. 410.778.2800
Fax 410.778.7275
andrei.straumanis@washcoll.edu

Larry Suter
National Science Foundation
Directorate for Education & Human Resources
4201 Wilson Boulevard, #855S
Arlington, VA 22230
Ph. 703.306.1655 x5820
Fax 703.306,0434
lsuter@nsf.gov

James Swartz
Grinnell College
1121 Park Street
Grinnell, IA 50112
Ph. 515.269.3100
Fax 515.269.4284
swartz@admin.grin.edu

Mary Swift
Biochemistry & Molecular Biology
College of Medicine
Howard University
Washington, DC 20059
Ph. 202.806.9749
Fax 202.806.5784
mswift@umd5.umd.edu

Hazel Symonette
Multicultural Affairs
University of Wisconsin–System Administration
780 Regent Street
Madison, WI 53715
Ph. 608.262.2275
Fax 608.263.4000
hsymonette@ccmail.uwsa.edu

Jorge Talamantes
Physics Department
California State University–Bakersfield
9001 Stockdale Highway
Bakersfield, CA 93311
Ph. 805.664.2335
Fax 805.664.2040
jtalamantes@csubak.edu

Richard Tapia
Rice University
6100 Main Street
Houston, TX 77005
Ph. 713.527.4049
Fax 713.527.4788
rtapia@rice.edu

Joe Teresa
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, DC 20208
Ph. 202.219.2046
Fax 202.219.2030
joe_teresa@ed.gov

Barbara Tewksbury
Geology Department
Hamilton College
198 College Hill Road
Clinton, NY 13323
Ph. 315.859.4713
Fax 315.859.4807
btewksbu@hamilton.edu

Barry Thompson
South Dakota State University
Box 507 Wenona Hall
Brookings, SD 57007
Ph. 605.688.4449
Fax 605.688.6074
thompsob@ur.sdstate.edu

Linda Tichenor
Biological Sciences
University of Arkansas
629 Science Engineering Building
Fayetteville, AR 72701
Ph. 501.575.6348
Fax 501.575.4010
tichenor@comp.uark.edu

Antoinette Torres
NACME
Programs
The Empire State Building
350 5th Avenue, Suite 2212
New York, NY 10118
Ph. 212.279.2626
Fax 212.629.5178
ttorres@nacme.org

Elias Toubassi
Mathematics Department
University of Arizona
617 N. Santa Rita, Bldg., #89
Tucson, AZ 85721
Ph. 520.621.6881
Fax 520.621.8322
elias@math.arizona.edu

Kenneth Travers
Office for Math, Science & Technology Education
University of Illinois at Urbana–Champaign
321 Armory Building
Champaign, IL 61820
Ph. 217.244.8286
Fax 217.333.7324

Jeff Turley
Arizona Collaborative for Excellence in the Preparation of Teachers
Physics & Astronomy
Arizona State University
Box 1504
Tempe, AZ 85287
Ph. 602.965.7907
Fax 602.727.6019
jtacept@asu.edu

Gil Valdez
North Central Regional Educational Laboratory
Center for Teaching, Learning & Curriculum
1900 Spring Road, Suite 300
Oak Brook, IL 60523
Ph. 630.218.1024
Fax 630.571.4716
valdez@ncrel.org

Ruth Vallejo
Science & Technology
University of Turabo-SUAGM
PO Box 3030
Gurabo, PR 00778
Ph. 787.743.7979 ext. 4113
Fax 787.744.5427

Kenneth Verosub
Geology Department
University of California
One Shields Avenue
Davis, CA 95616
Ph. 530.752.6911
Fax 530.752.0951
verosub@geology.ucdavis.edu

Carl Wamser
Chemistry Department
Portland State University
Portland, OR 97207-0751
Ph. 503.725.4261
Fax 503.725.3888
wamserc@pdx.edu

Deborah Warnaar
Chemistry Department
James Madison University
800 S. Main Street, MSC 7701
Harrisonburg, VA 22807
Ph. 540.568.7904
Fax 540.568.7938
warnaadl@jmu.edu

Norman Webb
National Institute for Science Education
1025 W. Johnson Street
Madison, WI 53706
Ph. 608.263.4287
Fax 608.263.6448
nlwebb@facstaff.wisc.edu

Robert Weinbeck
American Meteorological Society
AMS Education
1200 New York Avenue, NW, Suite 410
Washington, DC 20005
Ph. 202.682.9337
Fax 202.682.9341
weinbeck@dc.ametsoc.org

Dennis Weiss
Division of Science
City College of New York
Marshak Hall 1320
New York, NY 10031
Ph 212.650.6850
Fax 212.650.7948
dennis@scisvn.sci.ccny.cuny.edu

Laura Wenk
Natural Sciences
Hampshire College
West Street
Amherst, MA 01002
Ph. 413.582.5371
Fax 413.582.5448
lwenk@hampshire.edu

Gerry Wheeler
National Science Teachers Association
1840 Wilson Boulevard
Arlington, VA 22201
Ph. 703.312.9254
Fax 703.243.0407
gwheeler@nsta.org

Paula White
National Institute of Science Education
University of Wisconsin–Madison
1025 W. Johnson Street
Madison, WI 53706
Ph. 608.263.4353
Fax 608.263.6448
pwhite@macc.wisc.edu

John Wiley
University of Wisconsin–Madison
Bascom Hall, Room 150
500 Lincoln Drive
Madison, WI 53706
Ph. 608.262.1304
Fax 608.265.3324
wiley@macc.wisc.edu

Luther Williams
National Science Foundation
Directorate for Education & Human Resources
4201 Wilson Boulevard, #805N
Arlington, VA 22230
Ph. 703.306.1606
Fax 703.306.0399
lwilliam@nsf.gov

Mark Wood
Chemistry Department
Drury College
900 N. Benton
Springfield, MO 65802
Ph. 417.873.7474
Fax 417.873.7432
mwood@lib.drury.edu

Gordon Woodward
Mathematics & Statistics University of Nebraska–Lincoln
Lincoln, NE 68588-0323
Ph. 402.472.7239
Fax 402.472.8466
gwoodwar@math.unl.edu

Emmett Wright
College of Education
Kansas State University
1100 Mid-Campus Drive
237 Bluemont Hall
Manhattan, KS 66506
Ph. 785.532.7838
Fax 785.532.7304
birdhunt@coe.educ.ksu.edu

Lisa Wyatt
Charles Dana Center
University of Texas at Austin
ECN 3.200
2901 N. IH 35
Austin, TX 78722
Ph. 512.471.6190
Fax 512.471.6193
lwyatt@math.utexas.edu

Michael Zeilik
Physics & Astronomy
University of New Mexico
800 Yale Boulevard NE
Albuquerque, NM 87131-1156
Ph. 505.277.4442
Fax 505.277.1520
zeilik@chicoma.la.unm.edu

Appendix C
Acronyms

AAAS	American Association for the Advancement of Science
AAHE	American Association for Higher Education
ABET	Accreditation Board for Engineering and Technology
AMP	Alliance for Minority Participation (University of Puerto Rico)
CISE	Computer Information in Science and Engineering
DUE	Division of Undergraduate Education
EHR	Education and Human Resources
GPA	Grade Point Average
IHE	Institution of Higher Education
NISE	National Institute for Science Education
NSF	National Science Foundation
PI	Principal Investigator
REC	Research, Evaluation and Communication
SAT	Scholastic Assessment Tests
SMET	Science, Mathematics, Engineering, and Technology
SLL	Stanford Learning Laboratory
UPR	University of Puerto Rico
UT	University of Texas

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

X	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").