Tracking the Processes of Change in US Undergraduate Education in Science, Mathematics, Engineering, and Technology

ELAINE SEYMOUR
Bureau of Sociological Research, University of Colorado, Boulder, CO 80309, USA

Received 13 July 1999; revised 18 November 2000; accepted 14 December 2000

ABSTRACT: This paper describes some features in the changing landscape of activities intended to improve both quality and access in science, mathematics, engineering, and technology (SMET) undergraduate education. Observations are offered from the viewpoint afforded by my work—broadly over the last 10 years—both as a researcher, and as an evaluator for projects related to the improvement of undergraduate SMET education. Over that period, I have watched the landscape change—some issues, at first prominent, have diminished in importance; some are emergent; and yet others lie on the horizon. I have also observed that actions in pursuit of various reform goals reflect a variety of theories about how change can be accomplished that are not necessarily complementary. This short history of shifts in the focus of our efforts, and in our beliefs about how they may be achieved, is offered as a framework for discussion of these nationwide endeavors and as an aid in considering next steps. *Sci Ed* **86:**79–105, 2001.

SHIFTS IN THE LOCUS OF CONCERN

Over the last decade, the concerns that have driven efforts to improve the quality of SMET higher education in the United States have undergone a series of shifts. The first half of this paper reviews these transitions and some of the research that has informed them.

Correspondence to: Elaine Seymour; e-mail: seymour@spot.colorado.edu

Originally presented as a paper to the International Gordon Conference on Chemistry Education, Queen's College, Oxford, September 1998.

DOI 10.1002/sce.1044

"Pipeline" Issues: Focus on SMET Majors

Beginning in the mid-1980s, the work of the Higher Education Research Institute (HERI) at UCLA drew attention to a decline in the percentage of freshmen choosing to enter and remain in mathematics-and science-based majors. Their findings were based on longitudinal surveys of large, national samples of freshmen at 2- and 4-year institutions (Astin, 1985; Astin & Astin, 1993; Astin, Green, & Korn, 1987; Astin, Green, Korn, & Schalit, 1985; Dey, Astin, & Korn, 1991). In their 1993 report, Astin and Astin indicated that between freshman and senior years, science, mathematics, and engineering (SME)[1] majors suffered a relative student loss rate of 40%, largely, as Hilton and Lee (1988) pointed out, in the first 2.5 years of undergraduate work. Loss rates by discipline ranged from 50% in the biological sciences and 40% in engineering, to 20% in the physical sciences (when the transfer of former engineering majors into these disciplines is taken into account). Very few students transferred into SME majors after college enrollment, and there was always a net loss.

The highest losses occurred between high school graduation and college entry (NSF, 1990), leading to a steady decline in freshman enrollments in undergraduate science majors (from nearly 12% to just under 6%) in the 20 years prior to 1989, with an abrupt drop between 1983 and 1989 (Green, 1989a, 1989b), and declining enrollment in advanced degrees by American-born students (OTA, 1989). Initially, discussion of these findings led to concerns that there might be a shortfall in PhD-level scientists, mathematicians, and engineers by the turn of the century. Although these market projections were subsequently shown to be incorrect, the data on which they were based prompted concerns about the quality of education that was being offered to the next generation of scientists and engineers.

The Discovery of Underrepresentation in the Sciences

Losses from the pool of potential SMET majors came to be defined as "a problem" partly because we were also beginning to notice that the professions of science and engineering (and the student populations that supplied them) were "disproportionately" white and male. The HERI studies indicated a 20-year decline in women's SME enrollment (from almost 9% to just over 5%) despite enhanced recruitment efforts. Strenta and his colleagues (1993) further reported that the persistence rates of women entering SME majors were lower than those of men: male persistence rates varied between 61% for highly selective institutions and 39% for national samples, while the comparative rates for women ranged between 46% and 30%.

Astin and Astin (1993) also documented high loss rates among that smaller proportion of SME entrants that were Hispanic, African American, or Native American. Only one-third of Hispanics, one-half of African Americans, and one-half of Native Americans who enrolled in SME majors graduated in them. National data for the 1980s (Morrison & Williams, 1993) indicated that under 37% of students of color entering engineering programs completed degrees in that field, compared to over 68% for white students. Of those who continued into their sophomore year, about 57% of students of color graduated, compared with slightly more than 87% of white sophomores. Thus, the relative graduation rate for students of color in engineering was about half (52%) that of white students. The situation was no better in the sciences. Between freshman and junior years, 65% of students of color entering science or mathematics left their major, compared with 37% of white students (Culotta, 1992, p. 1209). Switching to non-SME majors accounted for only part of the loss: half of the students of color who left engineering dropped out of college altogether (Campbell, 1993).

[1] None of the early studies specifically included "technology" majors, and mention of this disciplinary group in research projects, issue papers, and the guidelines of funding agencies did not occur until about 1996. Before that date, "SME" rather than "SMET," is an appropriate descriptor. The addition represents yet another shift in focus.

Concerns about the causes and consequences of such underrepresentation generated debate about a further set of issues:

- Questions of inequity in educational and occupational access;
- Criticisms of the quality and character of the SME college experience and its role in underrepresentation;
- Workforce concerns: if the output from SME departments was limited because the pool of talent among young, white males was already well-exploited, then alternative sources of talent would need to be tapped.

The second and third concerns are connected. If SME college education was underperforming, then the losses among able white males could not be viewed as “normal attrition,” and the reservoirs of talent could not be regarded as either fixed or exhausted.

This set of concerns generated a considerable national effort to recruit more people of color into the sciences. By the early 1990’s, the National Science Foundation alone had spent over $1.5 billion in its effort to increase minority participation in science, and two major programs at the National Institutes of Health had invested $675 million in the same endeavor (Sims, 1992, p. 1185). In terms of recruitment, these initiatives were effective in dramatically increasing the enrollment of African Americans, Hispanics, and Native Americans. However, the outcome in terms of retention of students of color (other than Asian Americans) was very discouraging: although the number entering increased, the revolving door out of SME majors was spinning faster. By 1994, the enrollment of engineering freshmen of color had increased fivefold over that of the previous 20 years. Attrition rates, however, remained unchanged (cf. Brown, 1994, 1995; Massey, 1992; OTA, 1988). Across all SME majors, while white students had an attrition rate just over 27%, and Asian American students only 17%, about half of African American and Native American students, and two-thirds of Hispanic students left their SME majors (Astin & Astin, 1993).

As we argued from data from our study (Seymour & Hewitt, 1997), the main reason why retention efforts with women and students of color did not work better was because they supported individual students in an unremediated educational context. Precollege bridging programs, personal and academic support, and enrichment programs for underrepresented groups, targeted tutoring, and scholarships all have their place, but can only offer a temporary remedy. As our findings indicate, where such programs focus exclusively upon underrepresented groups and are not offered to all students, they may be avoided by targeted groups who see them as stigmatizing and create a backlash among the white, male majority. Programs targeting underrepresented groups also deflect attention from the greater challenge—which is to improve the quality of the undergraduate learning experience for all students. To paraphrase Einstein, you cannot resolve a problem in the conditions that created it.

The Loss of Able Students

Green (1989a, 1989b) also pointed out that the losses from undergraduate majors came from a pool of disproportionately able students. In 1988, slightly more than 45% of college entrants intending to enroll in SME majors had at least 10 semesters of mathematics and science, plus final high school GPAs of A or A- compared with 26% for students planning non-SME majors. Losses among women were of special concern, given their high overall ability: 60% of women enrolling in mathematics and the physical sciences, and 61% in engineering, had high school GPAs in the A range. On our own campus (the University of Colorado, Boulder), among freshmen entering between 1980 and 1988, women who chose SME majors had higher average Predicted GPA (PGPA) scores than their male peers

(i.e., 3.05 compared with 2.99 in engineering, and 2.84 compared with 2.72 in science and mathematics). Both women who persisted in the sciences, and those who switched to non–SME fields had higher average PGPA scores than men who either persisted or switched (McLelland, 1993).

Debate about the causes of underrepresentation in the sciences, and how best to address them, can, with hindsight, be seen as forcing the focus of interest in the reform of the sciences beyond its initial concern—largely with SME majors. This shift began with those researchers who:

- examined the factors affecting the choices of students in SME majors (e.g., Manis, Sloat, Thomas, & Davis, 1989);
- explored the nature of introductory science and mathematics classes as experienced by highly able students from other disciplines (Tobias, 1990), and directly challenged the (then) commonly held view that relatively few people are able to undertake science and mathematics courses;
- tested the possibilities of creating higher retention rates among students of color in these classes by changing their methods of teaching and creating more effective academic support mechanisms for students (e.g., Treisman, 1992).

This small body of work marked the beginnings of a concern that some aspects of the undergraduate SME experience might be contributing to failure to attract or retain able students, and that the pattern of losses might be (unwittingly) engineered rather than reflecting “natural” wastage.

To explore this proposition, we undertook a 4-year ethnographic study, whose aim was to discover, and to establish the relative importance of, those factors contributing to the decisions of SME and intended-SME majors at 4-year institutions to switch into disciplines that are not science-based. This work was funded by the Alfred P. Sloan Foundation. We interviewed only those students who had entered with mathematics SAT scores (or their equivalent) of 650 or above and who had declared (or had entered college intending to declare) SME majors. Approximately half of the sample had subsequently switched into non-SME majors, and half were graduating SME seniors. The interviews were undertaken at seven different types of institutions—four in Colorado and the rest in three different regions. Findings were validated by interviewing students on a further six campuses.

We discovered, somewhat to our surprise, that the same set of problems led both to switching and to a high volume of discontent with their educational experience among those who persisted (Seymour & Hewitt, 1997, pp. 30–40). We did not find switchers and nonswitchers to be two different kinds of people: they did not differ by performance, motivation, or study-related behavior to any degree that was sufficient to explain why one group left, and the other group stayed. Switching decisions proved never to be the result of a single, overwhelming concern, but were always the upshot of a “push and pull” process over time. The most common reasons for switching arose from a set of problems, which, to varying degree, were shared by switchers and nonswitchers alike. A far greater proportion of these difficulties arose from aspects of the undergraduate experience than from the inadequacies of individual students or from the appeal of non-SME disciplines. Paramount among these were reports of poor teaching, and difficulty in getting help with academic problems, which was mentioned by 90% of switchers and 74% of nonswitchers. Reports of inadequate high school mathematics and science preparation were offered by similar numbers of switchers (40%) and nonswitchers (38%).

What distinguished the survivors from those who left was not the nature of their problems, but whether they were able to surmount them quickly enough to survive. Serendipity also played a part in persistence, often in the form of intervention by faculty at a crisis point in

the student's academic or personal life. We found many switchers whose level of ability and application should have been sufficient—given a more encouraging learning environment—for them to complete their major. We also encountered a smaller number of multitalented switchers, the loss of whose high abilities from science-based fields may be of particular concern. On every campus, we also found a small group of "senior switchers" (nearly 17% overall) who were planning to leave their disciplines for non-SME careers following graduation.

The study sites were chosen to represent the types of 4-year institutions that most SME undergraduates attend. We presumed that the institutional context in which SME education takes place was likely to have some effect on retention and attrition. This presumption was not, however, supported by our findings. The concerns of both switchers and nonswitchers focused around the same set of issues across all seven campuses: every category of problem was found on every campus, regardless of differences in size, mission, funding, selectivity, or reputation. Although there were some variations in the ranking of problems by institutional type, there was little difference between the seven campuses in identification of the most serious concerns (Seymour & Hewitt, 1997, pp. 40–45).

We further analyzed our text data in terms of the differences in educational experiences—and their consequences—of women, and students of color from those of their white male peers. Students of color were distinguished by an array of different racial/ethnic backgrounds: any statement about "minority" students overall proved to be inaccurate and misleading. Some distinctive problems that contributed to the lower rates of persistence among each of these groups were identified. However, we also found that many problems leading to difficulties in persistence were shared with white male peers (Seymour & Hewitt, 1997).

The Quality of Mathematics and Science Education

In parallel with broader questions about the overall adequacy of precollege mathematics and science education, underrepresentation in the sciences also began to be discussed in terms of wide variations in the quality of preparation for college math and science preparation available to school children across the country. A pattern of gross inequalities by race/ethnicity, gender, location, school funding, staffing, facilities, and classroom resources was exacerbated by a chronic and growing shortage of discipline-qualified mathematics and science teachers in the K-12 system (Choy, Henke, Alt, Medrich, & Bobbitt, 1993; Dillworth, 1990; Henke, Choy, Geis, & Broughman, 1996; Horn, Hafner, & Owings, 1992; Schlechty & Vance, 1983; Weiss, Matti, & Smith, 1994).

Discussion of these issues extended far beyond academe. It was expressed in a public debate about the meaning and implications of studies indicating a poorer level of preparation in science and mathematics among school children in the United States compared with that in nations who are its economic competitors. In the most recent National Assessment of Educational Progress examinations in mathematics, only one in three students in grades 4 and 8, and 31% in grade 12 demonstrated basic competence, and 5% or less showed advanced competence (e.g., Reese, Miller, Mazzeo, & Dossey, 1997). In the Third International Mathematics and Science Study (TIMSS) undertaken in 46 countries (Schmidt, McKnight & Raizen, 1997), US students showed a steady decline in their mathematics and science performance between 4th and 12th grades—at which point, they ranked lowest in every category for both general and advanced levels of science and mathematics. Academic debates about the comparative methods employed by the TIMSS study (cf. Rotberg, 1998) have not lessened the level of public concern.

An additional concern was raised about variability in Advanced Placement (AP) courses by which students have traditionally prepared for college mathematics and science. Students from the United States with AP credits showed poorer conceptual understanding and ability

to solve problems than students with similar educational backgrounds in other countries (Juillerat, Dubowsky, Ridenour, McIntosh, & Caprio, 1997). In our (1997) study, many nonswitchers as well as switchers reported shock on discovering that their AP classes had not adequately prepared them for their first college mathematics and science classes.

The Shift from "Science-for-the-Few" to "Science-for-All"

This brings us to the main focus of current reform activity. Although efforts to improve the enrollment and retention patterns of underrepresented groups in SME—and now technology (T)—college majors continue, over the last 5 years, the primary thrust of reform activity has been to improve the science and mathematics competence of all students taking basic science and mathematics classes, and beyond that, schoolchildren and their teachers.[2] The presumption is, that a rising tide lifts all ships—although improvements in the general quality of college science and mathematics will benefit all students, they are argued to disproportionately benefit those who are poorly served by the existing undergraduate learning experience.

The reason for this shift is, as Mazur (1998) has observed, that competitive global market realities require that all educated citizens become science-and-math-literate. At an individual level, this will increasingly be required in order to achieve a good standard of living. The wider implications are both societal and moral: Tapia (1998) argues that SMET faculty have the collective power and opportunity to change the conditions that have created a permanent and growing underclass in US society, one cause of whose limited job options is lack of scientific, mathematical, or computing skills.

The vision of unfettered access to science is also a driving argument in the movement for virtual learning communities through Information Technology systems or Asynchronous Learning Networks. The most established of the "distance learning" institutions, the British Open University, has, since its inception in 1971, enrolled more than 2.5 million people in one or more courses (Van Dusen, 1997).

The target of reform is thus projected beyond the classroom to the department, the institution, and to all other articulating parts of the educational system. Its implication on campuses is that lower-level classes are an opportunity to discover which teaching strategies are most effective in delivering an adequate science and mathematics education to all college students. Using teaching and assessment methods as a means to discover "the few" by weeding out the rest is not only dysfunctional to this end, it is irrelevant.

This new vision was articulated in 1996 by two national reports—*From Analysis to Action* (NRC, 1996) and *Shaping the Future* (NSF, 1996), and by a further NRC report, *Transforming Undergraduate Education in Science, Mathematics, Engineering, and Technology*, in 1999. Taken together, these reports extend responsibility for the improvement of mathematics and science learning beyond those individuals, small groups, networks, consortia, and individual programs that have carried the burden of experimentation and the diffusion of innovation over the last decade, to everyone concerned with effective learning in mathematics and science at all levels. Because the states are "the only level of the educational system that can effectively and directly influence preschool through higher education, including the preparation of teachers,"[3] the National Science Foundation has approached the task of "systemic reform" by funding 23 state-based collaborations (State Systemic Initiatives) whose primary goal is to increase equity of access to K-12 mathematics and science education at a level sufficient for college preparation. Additionally, since 1994, it has targeted underserved

[2] This is reflected, for example, in the National Science Foundation's support for collaborative regional programs to improve teacher preparation in mathematics and the sciences.

[3] www.nsf.gov/StateSystemicInitiatives.

school populations through six Rural Systemic Initiatives, and, in 1999, consolidated earlier efforts to address the "continued disparity" between the academic performance of students in urban and suburban public schools by launching the Urban Systemic Program. This initiative supports partnerships of school districts that are implementing a standards-based science and mathematics curricula with 2- and 4-year colleges and universities. Whether these voluntary collaborations will prove robust enough to generate, and equitably distribute, resources sufficient to the task, and to build the infrastructure, professional commitments, and political will required to sustain comprehensive reform across the articulating parts of state and local educational systems is an open issue.

The three national reports also raise the need for accountability in the quality of SMET teaching (in both the K-12 and college sectors), and improvement in the quality of teacher preparation. As discussed earlier, other reports focus on the current shortage of discipline-educated mathematics and science teachers. The NSF has responded by supporting 20 Collaboratives for Excellence in Teacher Preparation. These regional alliances of institutions contributing to teacher preparation seek to recruit more students into mathematics and science education, increase their disciplinary knowledge by engaging SMET faculty more effectively in teacher preparation, and, in turn, infusing SMET faculty teaching with greater knowledge of research-based pedagogies.[4] The national reports stop short, however, of addressing the issue raised by President Clinton in his 1997 State of the Union address: "We should challenge more of our finest young people to consider teaching as a career." The greater contribution that SMET departments could make to the K-12 teaching force from among their own majors, and the career prospects that the public school system could offer them, are emergent issues awaiting wider national attention.

The Shift in Emphasis from Teaching to Learning

Our collective understanding of the task at hand has clearly been widening—from an initial, narrow focus on loss or wastage among SME undergraduate or graduate majors, to concern with the scientific and mathematical competence of potentially every child, student, and citizen. Those most intimately engaged with the implications of these shifts have become an interconnected nationwide community. They comprise a growing network of faculty experimenting with new modes of curriculum, pedagogy, and learning assessments in their own classrooms and departments, education researchers, program evaluators, and those public and private agencies who have promoted and funded innovation and the adaptation and dissemination of more effective SMET teaching methods. The fundamental value shift required in order to achieve "science for all" has emerged from the collective work of this large and growing group. Simply put, it is a shift in the focus of the classroom activities from teaching to learning. Its implications are not, however, simple. In classrooms, they include the following:

- refocusing classroom practice upon gains in student understanding, reasoning, application, and learning retention;
- clarification of student learning goals and their alignment with course assessments;
- redesigning assessments to engage students in their own learning and to give feedback to teachers on the efficacy of their work.

In and across SMET departments and their institutions, the implications include the following:

[4] See NSF (2000) for a discussion of these NSF programs, and of progress in establishing the National Digital Library.

- redefining, evaluating, and rewarding teaching and education scholarship as valued professional activities (Boyer, 1990; Glassick, Huber, & Maeroff, 1997);
- rethinking professional relationships—with colleagues in K-12 education, science education and research, assessment and evaluation research, other SMET disciplines, and in academe more broadly;
- restructuring professional education and development, including the primary professional education of graduate students, postdoctoral fellows and entering faculty, and reeducation for midcareer faculty;
- redesigning the facilities in which SMET courses are taught: learning-focused science education cannot be undertaken in facilities whose structures reflect an emphasis on passive learning (Project Kaleidoscope, 1996).

In my role as coevaluator for two linked Consortia for the Reform of Undergraduate Pedagogy (ChemLinks and ModularChem[5]), I have, over the last 5 years, tracked participating faculty's discovery of these, and other, ramifications of their collective undertaking. Although set in the context of a common agreement to focus on learning rather than teaching, early consortia discussions centered on the content (or "coverage") of particular new-style course offerings (referred to as "modules") that they were designing. Modules begin with questions, many of which have relevance for everyday life—concern about dietary fats, the causes and consequences of both global warming and the formation of the ozone hole, water purification, effective air bags, the utility of Vitamin C, and reasons for the annual reappearance of influenza. The module developers debated whether and how all the fundamentals of chemistry normally offered in introductory courses could be incorporated into the modular structure. Questions about what should be taught were quickly followed by discussions of appropriate teaching methods for the new materials. Most participants had experimented with alternatives to a formal lecture-demonstration-lab pedagogy, but their expertise and knowledge about how to teach more interactively varied. They taught themselves and each other, and attended workshops demonstrating "active" classroom strategies that complemented the inquiry-centered approach of their modules. They also discovered the need to assess the learning "gains" made by their students. At first, module developers and adapters used familiar assessment methods—tests, examinations, papers, lab reports, and projects—but quickly became aware that their assessments had much greater importance than merely ranking students. Assessments needed to be redesigned so as to give both faculty and students insights into the extent of learning gains—in understanding and skills, and in their ability to reason, formulate questions, make connections, apply their knowledge, and explain it to others.

Working out the practical implications of a new approach to content, pedagogy, and assessment methods takes time and experimentation. In interviews at a sample of 11 institutions where faculty were among the first to try out alpha versions of the modules, 67% of the students experiencing early modules highlighted insufficient "fit" between the various elements of the modular classes—class content, activities, text and web resources, and assessments. In a matched sample of more traditional classes in the same institutions 55% of the students also raised this issue (particularly lack of apparent fit between class and laboratory objectives), but the need for coherence is intensified where class structure and activities are less familiar.

[5] The two consortia have recently developed a combined identity as "ChemConnections" under which name they will continue their workshop and publishing activities beyond the ending of their original funding period as consortia.

The development and dissemination of new materials and methods, and their adaptation among faculty in funded collaborations and across more informal networks are underway in the physical, life and earth sciences, mathematics, and engineering. One of the ironies of the current situation is that what began out of concerns about poor student learning in the sciences has generated a growing body of faculty knowledgeable about learning theories and their practical application in these disciplines, while teaching methods remain largely unexamined in many nonscience classrooms. This said, the greatest single challenge to SMET pedagogical reform efforts remains the problem of whether and how large classes can be infused with more active and interactive learning methods. "Large" is a relative term. However, in our interviews with faculty and students at 21 institutions of six different types, we have found classes of about 80 students to be the largest number in which either direct dialogue between students and teacher, or group work integrated into the class structure, seems workable.

The common question, "What to do in a large class besides lecture?" has, after considerable experiment, produced a number of promising approaches. The simplest of these is to introduce into lectures short episodes of peer discussion, small group work on a question or problem, and other short activities that break up the session and engage students in actively understanding, applying, or extending class material. The chemistry coalitions (and other) web-sites[6] offer descriptions of these techniques. A second approach is to use computer technology, either to (re)design and equip the classroom so as to make live interaction possible, or to bypass large lectures altogether by placing all class materials (including hypertext and multimedia formats) on-line.[7] On-line courses may use student groups led by a teaching assistant "coach" as the main vehicle to gain mastery,[8] and offer minilectures on an "as needed" basis to clarify common areas of difficulty. Another technology-supported alternative is the "virtual classroom" that offers student–student and student–faculty computer teleconferencing as the delivery system, with computer-generated data and examples.[9] This also opens up the class to students learning at distance.

These two approaches require faculty control over the format of the classes that they teach. However, many university faculty are part of a departmental team that teaches large classes via one common lecture plus a number of laboratory and recitation sections. For these faculty, and others constrained by departmental norms to teach in traditional lecture mode, a middle path is to use recitation sessions run by teaching assistants as a way to insert more active learning, and checks on student comprehension, into an otherwise unchanged lecture and lab pedagogy. Success in this approach depends entirely on developing an effective method of training for TAs that ensures their active cooperation. This may not be easy, as faculty have found in developing a teaching assistant education program (based on worksheet activities) for recitation sections in a very large introductory chemistry class at the University of California, Berkeley.[10] Formative evaluation (Wiese et al., 2001) of the program indicates the strong attachment of many graduate student teachers to conservative teaching methods as an aspect of their preprofessional socialization. This problem is less pronounced where undergraduate TAs are used as peer leaders to augment lecture classes, as in the "Workshop Chemistry" program developed by the CUNY-based chemistry

[6] E.g., Project Kaleidosciope, *www.pkal.org*; NISE (collaborative learning pages) *www.wcer.wisc.edu/nise/cl1*.

[7] E.g., The CUPLE Physics studio at Renesselaer Polytechnic Institute (Wilson, 1994).

[8] Prof. Dick McRae's on-line introductory astronomy course at the University of Colorado, Boulder (Pedersen-Gallegos, Weston & Seymour, 2001, evaluation article forthcoming).

[9] E.g., Wilson & Mosher, 1994.

[10] Prof. A. M. Stacy and Dr. E. Lewis, Co-P.I. ChemConnections.

consortium, and now in use in over 50 institutions.[11] What is learned about graduate and undergraduate teaching assistant education, especially by faculty innovators working in research universities, will be of the utmost importance in meeting the challenge of improving student learning in more traditional large lecture classes.

The Centrality of Assessment to Higher Education Reform: New Forms of Classroom Assessment

The path that faculty active in the chemistry consortia have taken towards a shared understanding that assessment drives classroom reform is a journey increasingly shared by the whole faculty reform community. It has sometimes been described as "making a shift from valuing what we measure" to "measuring what we value." It begins with clarification of what exactly it is that you want students to learn. This then needs to be translated into corresponding classroom activities and assessments that allow both teachers and students to monitor gains in the processes of learning as well as their outcomes. Ebert-May (1998) has described this as measuring "active knowledge (understanding, reasoning, and utilization) rather than discrete, isolated bits of inert knowledge." Among faculty committed to learning-focused teaching, the design of better classroom assessments has come to be seen as the primary driver of change. It is feedback on what students actually learn that guides further adjustments in what is taught, and how.

In the first wave of innovation, faculty formed networks to share information about hitherto unfamiliar pedagogical techniques. Their current challenge is to discover, share, and test each other's assessment ideas. While working at the National Institute for Science Education (NISE), the University of Wisconsin, I helped to develop a web site, the Field-Tested Learning Assessment Guide (FLAG) to act as a repository, guide, and medium of exchange for classroom assessment methods generated by innovative SMET faculty for their own use, and offered for others to try. Much of the early material was donated by members of the "Establishing New Traditions" chemistry consortium based at the University of Wisconsin. Descriptions of assessment techniques (that now also include the published work of assessment specialists) are categorized by type; new materials will be continuously solicited and classified.[12]

Grass-roots initiatives of this kind have been one hallmark of the change movement. However, for such initiatives to prosper, they need long-term financial and technical support.

Rethinking the Evaluation of Teaching. Reconsideration of classroom assessment methods has been accompanied by a search for methods (for institutional as well as individual faculty use) that assess teacher efficacy in enabling learning. In addition to learning assessment instruments that match their revised learning objectives, classroom innovators have been seeking methods that:

- establish the student learning achievements of reformed teaching activities;
- offer a basis for argument and persuasion to promote further improvements in the quality of SMET education;
- protect faculty who work in contexts that are unsympathetic to classroom innovations from negative career consequences.

[11] www.sci.ccny.cuny.edu/∼chemwksp.

[12] The FLAG web site is maintained and further developed by the College Level One group of the National Institute for Science Education, the University of Wisconsin-Madison and can be found at http://www.wcer.wisc.edu/nise/cl1.

Institutions, disciplines, and national agencies are seeking data gathered at classroom, departmental, and institutional levels to develop aggregate measures of progress in improving the quality of SMET higher education. Researchers who work with national data sets referencing undergraduate education have, for some time, been lamenting their incompleteness. They are divided between those who see possibilities in the idea that common measures of student learning used across departments could be built into indicators of institutional progress, and those who see student learning gains and measures of institutional quality as lying in complementary, but different, domains.

At the department and individual faculty level, the issue is far from academic. Faculty teaching performance has, traditionally, been evaluated both, informally, by colleagues and, formally, by institutional or departmental end-of-semester classroom evaluation instruments. The criteria that govern faculty teaching reputations are largely implicit and are learned in the process of professional socialization. They are made explicit in classroom evaluation questions that focus upon class organization, lecture delivery, "coverage" of segments of the canon deemed appropriate for majors, and upon punctiliousness in carrying out formal duties. Many of the characteristics applauded in a good lecturer (sic) are those also recognized in a good performer—liveliness, reference to current research, and facility as a demonstrator. What institutional evaluation instruments typically fail to do is to explore and measure what students gain from various aspects of the class. For this kind of feedback, faculty often refer to students' write-in comments. Despite their inadequacy as sources of information on faculty efficacy as enablers of student learning, for tenure, or promotion decision purposes, classroom evaluation instruments are often the only available source of data about faculty as teachers (Cholakian, 1994; Tagomori & Bishop, 1994; Trout, 1997; Willliams & Ceci, 1997). Untenured faculty, or those who feel that their classroom innovations are unappreciated by departmental peers, need an alternative instrument that provides them with useful student feedback on the learning gains of their students in a form that can also protect the effective teacher from negative career consequences of classroom innovations.

I originally developed and field-tested the "Student Assessment of their Learning Gains" (SALG) instrument specifically in response to this need as expressed by faculty participating in the linked ChemLinks and ModularChem Consortia. The instrument is based on the hypothesis supported by our evaluation data that it is more relevant and productive to ask students what they have gained from specific aspects of the class than either what they liked or disliked about it, or to solicit judgements of their faculty as teachers. The SALG instrument avoids such questions and focuses exclusively on what students believe they have gained from particular aspects of their class that their teachers define as important. It can be used midsemester for corrective feedback, and also across two or more classes to track students' estimates of how well they were prepared for a current class by a prior class. With support from the Exxon-Mobil Educational Foundation, Sue Daffinrud (LEAD Center, the University of Wisconsin) placed the instrument on a web site where faculty can edit a template instrument to meet their individual requirements. It can be completed on-line by their students, and provides faculty with an instant statistical summary of the numeric data, and a print out of typed-in student comments. It may be found at http://www.wcwe.wisc.edu/salgains/instructor and is also available on the FLAG web site.[13]

[13] Accounts of the research underpinning the SALG instrument, and of two rounds of testing, can be found in reports to Exxon-Mobil Education Foundation (Seymour, Daffinrud, Wiese, & Hunter, 2000; Wiese, Seymour, & Hunter, 1999).

Indicators of Educational Quality at All System Levels. Nationally, the search is underway for forms of assessment that provide coherent, workable, and cost-effective indices of accountability for individual faculty, departments, and whole institutions. As I have argued earlier, such indices must be grounded, on the one hand, in more accurate and transferable measurement of student learning gains. On the other, they must be capable of assessing the effectiveness with which any institution, or the higher educational system overall, is providing quality SMET education for all students. We have not yet reached an understanding of how best to accomplish this. This search is also coupled with a call for reexamination of the criteria on which assessment instruments in common institutional use are based, and evaluation both of their utility and their consequences (Boyer, 1996; Tagomori & Bishop, 1995; Trout, 1997; Williams & Ceci, 1997).

THEORIES OF CHANGE; STRATEGIES FOR ACTION

Not only has our focus upon particular issues in SMET higher education shifted over time, so too have our ideas about how change may be accomplished. The strategies used to address the perceived shortcomings of SMET higher education reflect a set of diverse theories about how improvement may best be secured, and what conditions enable or constrain their chances of success.[14] In any field of endeavor, theory-building is a mental process by which we develop ideas that can allow us to explain why events should occur (Turner, 1982). Theories reflect the ways in which people engaged in any sphere of activity define that situation and decide what needs to be done: they generate and validate the strategies that the actors (in this case, educational reformers or innovators) see as the best ways to produce desired effects. However, in reform efforts, the theory or theories that underwrite the chosen forms of actions often remain unstated. Reformers may jump from identification of a problem to a selection of strategies intended to ameliorate it without reference to what is (or is not) known about the relative importance of the factors contributing to the problem, or about their chances of success. Their causal and predictive theories are, nevertheless, discernable in their chosen courses of action, and in the rhetoric by which they explain them. Embedded theories of reform are, thus, expressed in the ways reformers approach the innovations that they undertake. Collectively, they affect the direction of change—regardless of whether they may be considered "valid" by empiricists in the relevant disciplines.

The theories presented later are derived from three main sources: from my own research and evaluation work, and from ethnographic analysis of the written records of the 3rd Annual NISE Forum that are cited throughout this article. The 300 forum participants included a high proportion of those who are most actively engaged in efforts to improve the quality of science and mathematics education at all levels. They included faculty and administrators from 2- and 4-year institutions, teachers from the K-12 school system, educational researchers and evaluators, and representatives of public and private agencies that have promoted and funded innovations and their implementation. Some of the voices heard were of those who are deeply engaged in reform activities; some voices reflected a gradualist approach to change; others expressed degrees of skepticism about the need for change. All participants, however, were drawn to the forum by a common concern to find useful indicators of success by which to evaluate teaching and learning—whether innovative or traditional in nature—in classrooms, departments, and across institutions.

[14] Fullan (1999) reminds us that, because change is always complex, and the links between cause and effect are difficult to trace, a single theory of change is impossible, and a common theory of action is unlikely to be useful.

The records analyzed included, the speakers' papers; the tape-recorded and transcribed record of presentations, panel discussions, and audience contributions; and the written observations submitted by all participants as part of small group discussions following each panel. A digest of these records constitutes the proceedings of the forum (Millar, 1998). Access to the full record allowed me to code and classify the different types of action promoted by the speakers and writers, and to extrapolate the theories on which these choices of action appear to be based. What follows are the results of this analysis, augmented and illustrated by findings and observations from my current research and evaluation work.

One important theory of change—*that change is driven by shifts in what we value*—has been well illustrated in the first half of this paper. Other theories of change and their implications are discussed below.

Bottom–Up and Top–Down Theories of Change

Grass-roots and Network Theories of Change. Many activities to improve the quality of SMET higher education reflect the theory that:

reform across institutions or systems can be transmitted by the spread of grass-roots action between individuals, campus groups, and networks.

The difficulty with this (if used alone) is that it pays insufficient attention to the structural and cultural conditions within which networks operate. If the best of these practices is to thrive within institutions in the longer term, they will require support, recognition, and rewards from departments and institutions.

A second, more structured, version of "bottom-up" theories is 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.

The vitality and productivity of the participating ChemConnections faculty is clearly sustained by their conversational networks. In recognition of this, the consortia chose to spend a significant part of their funding on bringing subsets of the participants together for working meetings. Working out the details of new professional practice in the classroom in the course of such conversations is both essential (in situations of imperfect knowledge of what is required) and an intrinsic reward of involvement in innovation—particularly for more isolated faculty reformers. It is a source of intellectual stimulation, new learning and peer review. It sustains the innovators, especially those in situations of indifference and risk.[15]

National groups (especially funders such as the NSF, and resource centers such as the NISE) play an important role in supporting and extending grass-roots conversations and initiatives by providing workshops, working meetings and conferences (and/or travel money to attend them), funding for web-site maintenance, publicity, and recognition of individual or group achievements. They also require and fund program evaluations. Formative evaluation is especially important because the constant feedback it provides allows the group to become clearer about its values and goals, and suggests adjustments in its strategies (Fullan, 1999; Millar, 2000). What is not proven, however, is the theory that:

[15] Fullan (1999) argues that collaborative organizations generate the passion, energy, and commitment needed to pursue complex goals, and the emotional support needed to pursue them.

networks of such collaborations can build into a "critical mass" in favor of reform (Etzkowitz et al., 1995).

That there are some serious cultural barriers to change by grass-roots action is illustrated by the difficulties experienced by faculty who have attempted to spread educational innovations across disciplines. The members of the chemistry consortia periodically discussed the desirability of such collaborations, and on some campuses (often at the prompting of the NSF program officers or their National Visiting Committees) held cross-disciplinary gatherings. However, lacking a practical goal (such as curriculum articulation efforts that require cross-departmental negotiation and agreement), educational innovations developed in one department do not easily spread to another. The advent of educational technologies that require cross-departmental exploration of opportunities, costs, and some redistribution of resources or responsibilities may leverage greater cross-disciplinary collaboration than collegial networking unsupported by the prospects of rewards or consequences.

The dissemination efforts (and proposal requirements) of project funders often reflect a related theory that:

good ideas, supported by convincing evidence of efficacy, will spread "naturally"—that, on learning about the success of particular initiatives, others will become convinced enough to try them.

The evidence in support of this theory is also lacking. Indeed, there is some evidence that faculty do not respond to written accounts of positive findings for pedagogical experiments as they would to reports of research results within their disciplines. In the experiment described by Foertsch, Millar, Squire, and Gunter (1997), the research prestige both of the institution, and of the presenter, proved more persuasive than either the published evidence of the positive findings for a classroom innovation or a videotaped demonstration of its practicality, merits, and good reception by students.[16]

It is also not yet clear whether the classroom strategies generated out of the synergy created in faculty networks, such as the chemistry consortia, spreads to a finite or an expanding group of users. Interview data from faculty who are adapting the chemistry modules for use with their own classes and student populations indicate that the volunteer adapters are a self-selecting group. To date, they are mostly similarly-minded colleagues who already use various forms of contextual, interactive, collaborative, or discovery-based-learning in their classrooms who are open to trying out new methods. We have very few examples of "converts" to report.

In considering change strategies based on networking theories, it is important to discover why, for a variety of reasons, some faculty decide not to pursue their initial interest in any innovation and some reject it outright, and what contextual factors make it difficult for others to remain active. In order to clarify some of the limitations to change by networking, I have interviewed a sample of faculty who have been exposed to chemistry modules (by colleagues or through workshops), but who are not using the materials and methods they contain. The sample includes original consortia participants who never became fully engaged in the development and adaptation of modules or who became disengaged after an initial period of involvement, and faculty who attended workshops to learn about modules, but who subsequently decided not to use them, or used them only once. Analysis of these data is underway.

[16] Elmore (1995) goes further: "small groups of self-educated reformers apparently seldom influence their peers" (p. 20).

Value-Driven Institutional Leadership. Bottom–up approaches ultimately depend on the energy and organizational styles of natural leaders. In any collective human endeavor, energy and enthusiasm are cyclical, and require replenishment. Lacking a formal structure and renewable resources, grass-roots movements are apt to founder unless responsibility is taken for them by host institutions. Networks of faculty seeking change are obliged to negotiate with their departments, which hold a semiautonomous position within institutions. Although Oakes et al. (1998) and Fullan (1999) have argued that change that is driven solely from the top does not work, pressure from the institutional leadership can effectively legitimate and reward grass-roots reform activity. A number of SMET reformers have expressed doubt that change within higher education departments can occur without institutional leadership. As Gomez (1998) has argued, "Educational reform at the undergraduate level requires an institutional cultural transformation. Individual efforts of reform-oriented, proactive faculty are necessary, but not sufficient." Expressed as a theory of change:

system change within institutions requires unequivocal, high-level commitment to promote and reward classroom effectiveness and educational scholarship.

Achievement of either science-for-all, or learning-centered teaching are argued to require high-level institutional intervention to secure professional rewards for effective teachers and educational scholars. This, in turn, implies negotiation and implementation of institution-wide evaluation methods that reflect what students gain from their teachers' work.[17] This is also necessary to avoid reinforcement of the widespread cynicism with which faculty and many administrators have come to regard institutional evaluation activities. Institutions that follow this theory of change will need to plan for change at multiple levels, be ready to spend money and time to make it work, and to secure buy-in at the all-critical departmental level. They will need to identify, support, and recruit faculty who are leaders in educational improvement, and reinforce creative faculty collaborations. It would seem that "neither top–down, nor bottom–up strategies by themselves can achieve coherence" in reform efforts at the institutional level (Fullan, 1999, p. 27).

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

Regardless of where change begins, by this theory, faculty interested in new teaching methods cannot make progress without access to information about well-tested teaching and assessment methods.

Good intentions have to be channeled into actions that are already known to be effective. Time, effort, and resources cannot be wasted on strategies that have not worked well in other comparable settings.

The strategies to which this theory gives rise are professional development opportunities (including workshops) by which faculty and graduate students engaged in undergraduate teaching can build their teaching and assessment skills, and also gain an understanding of the learning theories on which they are based. Leadership at the institutional level is also required in order to sponsor and sustain professional development activities. This approach has been followed in the workshop model sponsored by the NSF (for example,

[17] Fullan (1999, p. 19) argues that, "policy initiatives that combine rigorous external accountability and mechanisms for focusing on local capacity development are critical for success."

in its summer workshops on teaching for new science and engineering faculty,[18] and also by private organizations, most notably, via the regional and discipline-specific workshops offered by Project Kaleidoscope.[19] An understanding that the self-education and unilateral innovation efforts of faculty are apt to founder without "buy-in" from more senior members of their institutions prompted Project Kaleidoscope to insist (from the outset) that all faculty attending their workshops must do so as part of a trio that also includes their department chair, and an administrator at the level of dean or above.

The resources needed both by newcomers and more experienced faculty innovators are as follows:

- access to pedagogical and assessment expertise, preferably local to their institution;
- teaching and learning assessment materials in accessible form;
- digests of pedagogical and assessment techniques;
- syntheses of the theoretical and research bases for these methods;
- evidence of their efficacy—including reports of what has not worked well.

The latter are not easy to find in conventional dissemination outlets. Reports of failed experiments are unlikely to reach peer-reviewed journals. Summaries of evidence about methods that do and do not work (plus their attendant caveats and conditions) have yet to be distilled from both published and unpublished research and evaluation reports. To meet this growing and unmet need, information that is generated by researchers and evaluators, largely in the fields of education and the social sciences, needs to be made available in a form and linguistic style that is accessible to SMET faculty. Developing a distilled and annotated bank of information will require commissions to writer–editors, access in both print and electronic forms, funding for continuous collection and classification, and maintenance as a permanent electronic resource. This has been one function proposed for the National SMET Digital Library. Other examples of national resources currently being developed to meet these needs are the NSF's Online Evaluation Resource Library, and individual consortia web sites. The NSF has also begun to build evaluation strategies that will collect data across initiatives of similar types. However, in order to provide a better supply of quality project evaluation, the funders of educational initiatives might also be more specific and demanding in their guidelines for required evaluation and dissemination components. This would involve funding agency staff in a proactive enabling function:

- helping program principal investigators (PIs) find evaluators who are appropriate to the project and, preferably, local;
- publishing guidelines on how to work with evaluators (e.g., approaching an evaluator early in the planning stage, negotiating a project-appropriate strategy, grounding the evaluation budget in the work required);
- smaller projects may need more direct assistance in developing and implementing appropriate evaluation strategies.

To raise the quality of program evaluations and ensure that relevant findings inform new proposals requires the availability of clear, synthesized information about findings from prior programs, about the theoretical underpinnings of commonly proposed strategies, and

[18] E.g., the NSF Engineering Education Scholars' Program.

[19] For example, the Faculty for the 21st Century workshops, Project Kaleidoscope, The Independent Colleges Office, Washington, D.C.

guidance in their use. It also requires that funders take a more active role in enabling PIs to fulfill their evaluation requirements.

One source of "blueprints" that has been greatly underused by SMET classroom innovators is knowledge about the nature of learning, and expertise in curriculum development, pedagogy, and learning assessment techniques that is to be found in Colleges of Education. This resource is available to SMET faculty and administrators at many 4-year institutions, and lies in reasonable proximity to others. That collaborative relationships are not, as yet, much explored reflects an unfortunate history that includes the devaluing by SMET faculty of teaching—as a professional activity, as a career for their students, and as a form of scholarship (Seymour & Hewitt, 1997). The consequence is a record of limited contact between SMET faculty and science and mathematics education specialists. This situation may be improved by the NSF's "Collaboratives for Excellence in Teacher Preparation," and by the State Systemic Initiatives, both of which bring together representatives from these two faculty groups with master teachers from the K-12 system in order to address the national crisis in the quality and supply of K-12 science and mathematics teachers. They offer one practical model by which to break stereotypes and the neglect of a valuable national resource.

Alignment is Required at All Levels for Effective System Change

As it became clearer that system-wide changes would be called for to achieve the goal of science-for-all, innovators began to recognize the need for various forms of alignment—a need that had already been articulated and implemented as part of K-12 system reforms. In higher education, alignment is seen as desirable at a number of levels—progress in each of which is, as yet, partial. Its dimensions are alignment of the following:

- classroom assessment practices with student learning goals;
- 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;
- curriculum developments in SMET higher education with developments in K-12 reforms, with the ultimate goal of an educational continuum;
- course offerings across departments to produce cross-disciplinary degrees. (This form of alignment may be a response to external pressures, especially new market opportunities for graduates);
- the activities of SMET faculty classroom reforms with the knowledge and skills of colleagues working in the disciplines of education, assessment, and evaluation;
- data collection practices at the national level such that national data sets can better inform evaluation practices at the institutional level;
- improved classroom evaluation methods in order to meet departmental and institutional needs for evidence of teacher effectiveness, and faculty need to receive meaningful feedback about student learning gains.

The most basic form of alignment for individual faculty occurs at the classroom level. Expressed as a theory:

In order to make the curriculum more meaningful to students, faculty must articulate their learning goals, align their teaching and assessment strategies with these goals, and make students aware of their own learning processes.

Student interview data from the modular chemistry classes on 11 campuses supports this theory: the need for alignment of important course elements (which students referred to as matters of good “fit”) was raised spontaneously in 67% of the modular student interviews. Students looked for alignment between class content and class assignments, between class and lab learning objectives, and between the order of content presented in class and that of their texts. They wished to see these connections clearly and openly explained. Such concerns are likely to be found among students in innovative classes partly because the relationship between class elements will be less familiar to students, and partly because an experimental class may still be “under construction.” However, the issue was also strongly expressed as an unmet need in 55% of the interviews with students in the matched traditional classes. In those classes, the most commonly expressed concern was lack of understanding of the connection between the content of lectures and the purpose of (presumably related) labs. Taken together, these student observations endorse the theory that:

Learning is enhanced when all of the main elements in a class fit coherently and overtly together: class content and activities, lab work, assignments, the text, media, and other resources.

Students especially wanted to see connections between unfamiliar class elements “sign-posted.”

Alignment also makes a de facto system coherent. Unlike the formally aligned educational systems of many other countries (where the linkages are planned and orchestrated at a national or regional level), a system that is built from a convergent, but local and independent basis, cannot take coherence for granted. Such a system is also more difficult to change, because to make any progress requires that we identify the most powerful elements in the system, and then figure out how to use them to create leverage. A concern with alignment thus reflects the theory that:

attempts to alter single elements in a complex social system will not be effective: each element must be aligned with the others for system changes to prevail.

Departmental Values are Key to Educational Improvements

As suggested earlier in this analysis, the department is the rock against which the teaching innovations of individuals or small groups of faculty are most apt to founder. Departments operationally define, structure, evaluate, and reward the teaching and learning activities of higher education institutions. The theory to which this definition of the situation gives rise is as follows:

Finding the means to leverage relevant shifts in departmental values and practices is the critical factor in determining whether the efforts of faculty—as individuals and groups—and of their institutions, will be able to improve the quality of SMET education, or achieve the wider goal of science-for-all.

As also discussed earlier, teaching effectiveness is traditionally evaluated by peer reputation and student classroom evaluation. The criteria by which departments judge faculty as teachers can be observed both in these instruments, and in collegial arguments about curriculum change. Departmental values traditionally include the primacy of research over teaching, a primary loyalty to the discipline over the institution, and a duty as teachers

to "cover" collectively approved segments of the canon. The standards and practices that reflect these priorities, as Merton (1942) originally observed, transcend particular nations or institutions. Practices that are interpreted as serving the universal standards of science are hard to question or to amend. The primary departmental defense against the arguments of those colleagues who have already refocused on increased student learning and the goal of "science-for-all" is the argument that universal disciplinary standards must prevail.

The Principal Investigator of the ChemLinks Consortium, Brock Spencer, illustrated this defense in action when describing the concern of some colleagues that his "lectureless" approach to teaching undergraduate chemistry (in which much of the learning is done in small groups) "lacked rigor." Colleagues cited the drop-out 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%). Persuading colleagues to see their primary teaching goal as better student learning, their classroom activities as an open-ended experiment, and goal-assessment criteria alignment as the primary task involves a paradigm shift. The question for the reform community is, How can this attitude shift happen?

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

Among the community of faculty classroom innovators, program evaluators, and funders there is a gathering consensus that, because both validation and tangible rewards shape faculty goals and actions,

> the fastest and most enduring way to promote a renewed emphasis on teaching in the service of learning in higher education is to restructure the faculty rewards system.

As Tapia (1998, pp. 16–17) has asserted, rewards "drive the whole system," and "are determined at the departmental level." Not only is review of the rewards structure seen as critical in promoting improvements in the quality of higher education teaching, it is necessary to protect the current generation of innovators, better secure the institutionalization and spread of successful innovations, and allow faculty to choose classroom scholarship as a professional focus.

Changing the criteria for departmental rewards—tenure, promotion, resources, time, and opportunities for professional development—will, again, require leadership. Leverage exercised by powerful groups will be required to even the imbalance between rewards for teaching and education scholarship, on the one hand, and for discipline-focused research on the other. Accreditation agencies, public and private funding agencies, and national testing services are some of the external bodies with the power to influence the direction of departmental values and priorities. Leadership by presidents, provosts, and deans will also be required—although the record of effective departmental resistance to the implementation of institutional goals, plans, and exhortations is also acknowledged.

The consequences of failing to find effective means to address this barrier to reform are casualties among the current generation of creative teachers, and loss to the education reform effort of the next wave of faculty—through pedagogical knowledge not gained, and socialization into prevailing departmental attitudes and practices while waiting for the safety of tenure. Among faculty active in development and testing of the chemistry modules, two were denied tenure, and a third was awarded tenure after 1 year of probation and review. As communicated to these three faculty, collegial judgements were influenced by greater experiment and scholastic productivity in education than that in research. Tenured faculty who are active in educational innovation often express reluctance to encourage untenured

colleagues in their departments to devote any significant amount of their time to classroom experimentation or educational scholarship. Speakers at the 1998 NISE Assessment Forum unanimously (though with regret) advised against the engagement of untenured faculty in educational innovation and scholarship: as Tapia (1998) observed, “faculty cannot change the system unless they first survive it.”

Balancing respect for educational scholarship with that for disciplinary research in departmental reward systems necessarily involves assessment of student learning gains as a measure of teaching competence, and recognition of contributions to peer-reviewed education scholarship. This is unlikely to occur unless the criteria for rewarding teaching and educational scholarship are aligned with existing departmental values. Boyer (1990) and other scholars (Diamond & Bronwyn, 1995; Edgerton, 1995; Glassick, Huber & Maeroff, 1997; Seldin, 1989) have argued that the traditional criterion by which research is judged should also be applied to teaching and educational scholarship—that is, as peer-recognized intellectual work, appropriately disseminated.

Evidence is a Necessary (if not Sufficient) Condition for Reform

In order to promote and validate such an adjustment in the criteria for faculty rewards, or to diffuse successful innovations, reformers propose that:

it is necessary to provide clear and convincing evidence that innovative forms of teaching are as effective as, or more effective than, traditional approaches to teaching. It is not enough to claim that greater learning occurs; it must be demonstrated.

Faculty innovators argue from their experiences in promoting and defending their innovations with colleagues that failure to supply these proofs will be an absolute barrier to change. Colleagues demand evidence of comparable or improved student learning, and administrators want data on the effectiveness of faculty performance.

For a number of reasons, meeting this requirement is not a simple matter. As the chemistry module developers have found, new approaches to content or pedagogy require a period of classroom adjustment. There may be initial difficulties in making the elements of a new pedagogy fit coherently together; too much may be attempted, and content or activities have to be scaled back; unforeseen practical problems arise (with labs, software, materials, etc.); and faculty may feel awkward using less familiar teaching methods for a while.

While teachers are in the midst of resolving these difficulties, they are likely to encounter resistance from students who have learned how to get good grades by more passive learning methods and who find it harder to achieve the level of grades they have come to expect by a pedagogy that often demands more of them.[20] As our evaluation data indicate, the degree of student resistance varies by class size, institutional character, student population, and by the dominant attitudes of students, faculty, and teaching assistants. The chemistry innovators have found greater student dissent in the very large classes at competitive research universities where both colleagues and graduate students are strongly socialized into traditional methods. As the module adapter interview data reveal, student resistance occurs in other settings. However, it has proved easier to gain acceptance for and appreciation of more active forms of learning in classes under 100, in classes at liberal arts and 2-year colleges, and in two-semester (or term) sequences where students have time to reconsider and adjust their learning habits. We note, however, that where students receive indications that they

[20] Indeed, both Maurer (1996) and Fullan have argued, “resistance is an essential ingredient of progress” (1999, p. 23).

have learned well (e.g., by their grades, their ability to understand the material in journal articles, to interpret data, explain material to others, or discover that they are well prepared for a subsequent class) they may still question if they can "know" something that they did not memorize, or whether enjoying a class means that they cannot have learned as much chemistry. This is compounded among freshmen who cannot (as yet) distinguish between the adjustments required to close the gap between high school and college science and the discomforts peculiar to modular (or other) innovative forms of teaching.

All of these are likely to affect the classroom evaluation scores students give their teachers. This leaves innovative teachers vulnerable until they gain proficiency with the new methods and students begin to recognize their benefits. It again underscores the need for forms of classroom evaluation that reflect what students have gained from their class rather than what they "liked" about the class or their teacher.

Although proof of efficacy for well-developed classroom innovations is needed for them to be taken seriously, this does not, in and of itself, prompt colleagues to take note of the evidence or to see it as a valid reason to change in their own professional practice. Shifts in scientific theory do not occur as an automatic response to accumulations of data (Kuhn, 1970). When the shift that is called for is a shift in values and social behavior (rather than in disciplinary thought and practice), the response, as Tobias (1992) has observed, is often unaffected by available evidence. Brock Spencer has illustrated this point from experience: although his chemistry colleagues acknowledge that the students from his lecture-less class do at least as well in subsequent classes as those who elect more traditional sections, they may still not be convinced by the evidence. This common experience among innovators is supported by the research (e.g., Foertsch et al., 1997; Nonaka & Takeuchi, 1995), indicating that the personal endorsement of classroom innovations by colleagues who are esteemed for their research, and/or for their institutional prestige, is more important than reading or hearing accounts of evidence. Although there are no guarantees that good proofs of enhanced student learning—though necessary—will be sufficient to convince skeptics to change their classroom practice, adding three caveats to the theory may increase the chances of change by dissemination of evidence:

- skeptics and interested inquirers alike must be able to make sense of the data;
- the data must illuminate outcomes that are valued by the targeted audience;
- personal endorsement of new teaching practices, and/or the evidence in their favor, by colleagues with high research prestige may be required.

For example, an indicator whose value is appreciated by faculty and administrators of all pedagogical persuasions is evidence of the levels of understanding and skills that students carry forward into another class in a disciplinary sequence, or carry away into other disciplines or into the "real world." Like other evaluators, we are experimenting with ways to demonstrate student learning retention.

The proving imperative also raises the question of why innovators are required to provide evidence of the efficacy of their methods when we lack good evidence of the efficacy of more traditional forms of teaching (Mazur, 1997). As Lawson (1995) observes, "remarkably little of what is introduced in text and lecture is understood and retained." Indeed, innovation is often prompted by such concerns, and by doubts that the array of tests, examinations, reports, and projects in common use are reliable indicators of learning. Reformers tend to be held to higher and different standards of proof than more traditional practitioners, and evaluators, thus, lack baseline data against which to make valid comparisons.

Neither can assessments developed for one method of instruction be used to measure student learning gained by other methods. When instruction goals and methods are changed,

their impact on learning can only be measured by a corresponding change in assessment methods. The assessment scores of classes with different goals, pedagogies, and types of assessments cannot be meaningfully compared. As Mazur (1998, p. 93) has argued, "Changing the methods of assessment means giving up any meaningful correlation with previous assessments. If administrators and faculty do not realize that this poor correlation is an unavoidable consequence of change, it will be impossible to move forward."

In light of these arguments, the proving imperative might be rewritten thus:

it is necessary to provide clear and convincing evidence that all forms of teaching (whether "innovative" or "traditional") are effective in promoting student learning. It is not enough to claim that learning occurs; it must be demonstrated.

Change by Leverage from External Agencies

Finally, the success of the Accreditation Board for Engineering and Technology (ABET) in promoting nationwide changes in undergraduate engineering curriculum and pedagogy has encouraged the hope that:

change may be leveraged by agencies external to institutions.

Accreditation agencies, the national testing services, scientific societies, the National Academy, and public and private funding agencies may have individual and collective power to leverage change through their direct influence on institutions—most especially, on departments.

This is a new role for accreditation and national testing agencies, which, historically, have been seen as exerting a conservative influence on the curriculum. Indeed, many SMET cite faculty accreditation criteria, or the presumed, conservative character of professional examinations as either a partial or absolute barrier to change. Members of the Chemistry consortia report a shift towards exploration of students' conceptual grasp and ability to apply knowledge in the types of questions asked in the MCAT examinations. If this change is more widely noted, it may allow those departments that "service" large numbers of premedical students to reconsider aspects of their curriculum. The scientific societies, however, are still viewed as exerting a conservative influence on curriculum development, classroom assessment practices, and cross-disciplinary collaboration.

One long-standing model of change that makes use of external agency leverage is grants-driven reform. Groups of innovators are funded to develop and test an educational initiative on the premise that:

the time for development, implementation, and testing that agency grants provide, plus the prestige of such awards, will increase the chances that innovation will take root in the host institutions beyond the end of funding.

This "priming the pump" approach to change has been an important strategy for both public and private foundations for at least two decades. Outside funding has been critical in developing and testing many educational experiments. However, it may be less effective in securing the institutionalization of innovations whose worth has already been demonstrated. One unfortunate side-effect of the system of funding for educational experimentation is that it follows the model used for competitive research funding. As Coppola (1998) has argued, it is a system that 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, in turn, reduces the chances for changing departmental values, and, thus, for the survival of the initiative—unless, that is, host institutions undertake that responsibility.

What is required is more funding for the following:

- departments, institutions, or collaboratives that adapt models with an established record;
- higher education research within the SMET disciplines;
- curriculum development projects that build upon such research;
- whole-organization projects where institutional buy-in (including matching funds and resources) is required from the outset.

NEXT STEPS

Action is now beginning to be directed towards clarification of what “science-for-all” might look like. Its implications clearly include better alignment of the efforts of science and mathematics teachers in the K-12 and 2-year colleges, with redesigned introductory SMET classes in the 4-year system. The analysis must include the costs of change, what reallocation of resources within and between institutions will be required, how those may be achieved, and the implications of a changing job market. The new questions for the research and evaluation community and the education professionals will be: How shall we measure faculty efficacy both in their teaching role and as educational scholars? How can we best reward these professional activities? How will faculty develop the skills they need to do them well? Most fundamentally, we must find better answers to the question: How can we increase the supply, and raise the quality of, the mathematics and science teaching force—which is the foundation on which the vision of science-for-all ultimately rests?

The author acknowledges the University of Colorado, Boulder, and John Wiley and Sons, Inc., publishers, for their support of this work in progress; the members of the ChemLinks and ModularChem Chemistry Consortia, and the National Institute of Science Education, University of Wisconsin, Madison, for access to the records of their Third Annual Forum, *Indicators of Success in Post-Secondary SMET Education: Shapes of the Future*, February 1998.

REFERENCES

Astin, A. W. (1985). Achieving educational excellence: A critical assessment of priorities in higher education. San Francisco: Jossey-Bass.

Astin, A. W., & Astin, H. S. (1993). Undergraduate science education: The impact of different college environments on the educational pipeline in the sciences. Los Angeles, CA: Higher Education Research Institute, UCLA.

Astin, A. W., Green, K. C., & Korn, W. S. (1987). The American freshman: Twenty-year trends, 1966–1985. Los Angeles, CA: Higher Education Research Institute, UCLA.

Astin, A. W., Green, K. C., Korn, W. S., & Schalit, M. (1985). The American freshman national norms for fall, 1985. Los Angeles, CA: Higher Education Research Institute, UCLA.

Boyer, E. (1990). Scholarship reconsidered: Priorities of the professoriate. Report of the Carnegie Foundation for the Advancement of Teaching. San Francisco: Jossey-Bass.

Boyer, E. (1996). From scholarship reconsidered to scholarship assessed. QUEST, 48, 129–139.

Brown, S. V. (1994). Under-represented minority women in science and engineering education. Princeton, NJ: Educational Testing Service.

Brown, S. V. (1995). Profiles and persistence of minority doctorate recipients. Final report to the Graduate Records Examination Board. Princeton, NJ: Educational Testing Service.

Campbell, G., Jr. (1993, March 31). Visions of engineering education in century II. The Porth Distinguished Lecture, University of Missouri at Rolla, MO.
Cholakian, R. (1994). The value of evaluating. Academe, 80(5), 24–26.
Choy, S. P., Henke, R. R., Alt, M. N., Medrich, E. A., & Bobbitt, S. A. (1993). Schools and staffing in the United States: A statistical profile, 1990–91 (NCES 93–146). Washington, DC: National Center for Education Statistics.
Coppola, Brian P. 1998. Assessment and the promotion of change in departments, disciplines, and institutions: The reaction to the symptoms versus reaction to the disease. Paper and presentation to the National Institute for Science Education Forum. Indicators of success in post-secondary SME&T education: Shapes of the future, February 23–24. In S. B. Millar (Ed). 1998. Synthesis and Proceedings of the Third Annual NISE Forum, University of Wisconsin-Madison, NISE, WCER, pp. 101–112.
Culotta, E. (1992). Scientists of the future: Jumping high hurdles. Science, 258(5085), 1209–1213.
Dey, E. L., Astin, A. W., & Korn, W. S. (1991). The American freshman: Twenty-five year trends. Los Angeles, CA: Higher Education Research Institute, UCLA.
Diamond, R. M., & Bronwyn, E. A. (Eds.). (1995). The disciplines speak: Rewarding the scholarly, professional and creative work of faculty. Washington, DC: AAHE.
Dillworth, M. E. (1990). Reading between the lines: Teachers and their racial/ethnic cultures. Washington, DC: Clearinghouse on Teacher Education.
Ebert-May, D. 1998. Assessment as a learning process: What evidence will we accept that students have learned? Paper and presentation to the National Institute for Science Education Forum. Indicators of success in post-secondary SME&T education: Shapes of the future, February 23–24, 1998. In Millar S. B. (Ed). 1998. Synthesis and Proceedings of the Third Annual NISE Forum, University of Wisconsin-Madison, NISE, WCER, pp. 5–11, 79–89.
Edgerton, R., (1995 April 7–8). Evaluating teaching as scholarly work. Keynote address, Academic Affairs Symposium, University of Georgia, Athens, GA.
Elmore, R. (1995). Getting to scale with good educational practice. Harvard Educational Review, 66(1), 1–26.
Etzkowitz, H., Kemelgor, C., Neuschatz, M., Uzzi, B., & Alonzo, J. (1995). Gender implosion: The paradox of "critical mass" for women in science. Paper to the NACME Research and Policy Conference on Minorities in Mathematics, Science and Engineering, Wake Forest University, Wake Forest, NC.
Foertsch, J. M., Millar, S. B., Squire, L., & Gunter, R. (1997). Persuading professors: A study in the dissemination of educational reform in research institutions. Report to the NSF Education and Human Resources Directorate, Division of Research, Evaluation, and Communication, Washington DC. Madison: University of Wisconsin-Madison, LEAD Center.
Fullan, M. (1999). Change forces: The sequel. Philadelphia, PA: Falmer Press.
Glassick, C. E., Huber, M. T., & Maeroff, G. I. (1997). Scholarship assessed: Evaluation of the professoriate. An Earnest Boyer Project of the Carnegie Foundation for the Advancement of Teaching. San Francisco: Jossey Bass.
Green, K. C. (1989a). A profile of undergraduates in the sciences. The American Scientist, 78, 475–480.
Green K. C. (1989b). Keynote address: A profile of undergraduates in the sciences. In An exploration of the nature and quality of undergraduate education in science, mathematics and engineering, national advisory group, Sigma Xi, the scientific research society. Racine, WI: Report of the Wingspread Committee.
Gomez, M. 1998. An assessment model to drive undergraduate educational reform in SMET fields in a large public, multi-campus university system. Paper and presentation to the National Institute for Science Education Forum, Indicators of success in post-secondary SME&T education: Shapes of the future, February 23–24. In S. B. Millar (Ed) 1998. Synthesis and Proceedings of the Third Annual NISE Forum, University of Wisconsin-Madison, NISE, WCER, pp. 23–29, 1333–1136.
Henke, R. R., Choy, S. P., Geis, S., & Broughman, S. P. (1996). Schools and staffing in the United States: A statistical profile, 1993–94 (NCES 96-124). Washington, DC: National Center for Education Statistics.

Hilton, T. L., & Lee, V. E. (1988). Student interest and persistence in science: Changes in the educational pipeline in the last decade. Journal of Higher Education, 59, 510–526.

Horn, L., Hafner, A., & Owings, J. (1992). A profile of American eighth-grade mathematics and science instruction (NCES 92-486). Washington, DC: National Center for Education Statistics.

Juillerat, F., Dubowsky, N., Ridenour, N. V., McIntosh, W. J., & Caprio, M. W. (1997). Advanced placement science courses: High school—college articulation issues. Journal of College Science Teaching, 27, 48–52.

Kuhn, T. S. (1970). The structure of scientific revolutions. Chicago: University of Chicago Press.

Lawson, A. (1995). Science teaching and the development of thinking. Belmont, CA: Wadsworth.

Manis, J. M., Sloat, B. F., Thomas, N. G., & Davis. C. (1989). An analysis of factors affecting choices of majors in science, mathematics and engineering at the University of Michigan. University of Michigan, MI: Center for Continuing Education for Women.

Massey, W. (1992). A success story amid decades of disappointment. Science, 258(5085), 1177–1179.

Maurer, R. (1996). Beyond the wall of resistance. Austin, TX: Bard Books.

Mazur, E. (1997). Peer instruction. Upper Saddle River, NJ: Prentice-Hall.

Mazur, E. (1998). Moving the mountain: Impediments to change. Paper and presentation to the National Institute for Science Education Forum, Indicators of success in post-secondary SME&T education: Shapes of the future, February 23–24, 1998. In Millar, S. B. (Ed). 1998. Synthesis and proceedings of the Third Annual NISE Forum, University of Wisconsin-Madison, NISE, WCER, pp. 5–11, 91–93.

McLelland, L. (1993). Students entering science, mathematics, and engineering majors as fall freshmen, 1980–1988. Unpublished data provided by L. McLelland, University of Colorado, Boulder, Office of Research and Information.

Merton, R. (1942, October). Science and technology in a democratic order. Journal of Legal and Political Sociology, 1, 115–126.

Millar, S. B. (Ed). (1998). Synthesis and proceedings of the third annual NISE forum. University of Wisconsin-Madison, NISE, WCER.

Millar, S. B. (2000). *The role of formative evaluation in the development of an inter-disciplinary academic center* (NISE Occasional Papers #8). Madison, WI: Wisconsin Center for Educational Research.

Morrison, C., & Williams. L. E. (1993). Minority engineering programs: A case for institutional support. NACME Research Letter, 4(1), 156–167.

National Research Council. (1996). From analysis to action: Undergraduate education in science, mathematics, engineering, and technology. Report of a Convocation. Washington, DC: National Academy Press.

National Research Council. (1999). Transforming undergraduate education in science, mathematics, engineering, and technology. Committee on Undergraduate Science Education, Center for Science, Mathematics, and Engineering Education. Washington, DC: National Academy Press.

National Science Foundation. (1990). The state of academic science and engineering. Directorate for Science, Technology, and International Affairs, Division of Policy Research and Analysis. Washington, DC: NSF.

National Science Foundation. (1996). Shaping the future: New expectations for undergraduate education in science, mathematics, engineering and technology (NSF 96-139). Washington, DC: National Science Foundation.

National Science Foundation. (2000). Annual report, division of undergraduate education, directorate for education and human resources, fiscal year. 2000. On-line: www.nsf.gov.

Nonaka, I., & Takeuchi, H. (1995). The knowledge-creating company. Oxford: Oxford University Press.

Oakes, J., Welner, K., Yonezawn, S., & Allen, R. (1998). Norms and policies of equity-minded change. In A. Hargreaves, A. Lieberman, M. Fullan, & D. Hopkins (Eds.), International Handbook of Educational Change (pp. 952–973). Dondrecht: Kluwer,

Office of Technology Assessment. (1988). Educating scientists and engineers: Grade school to grad school. Washington, DC: GPO.

Office of Technology Assessment (1989). Higher education for science and engineering: Background paper. Washington, DC: GPO.

Project Kaleidoscope (1991). What works: Building natural science communities; A plan for strengthening undergraduate science and mathematics (Vol. 1). Washington, DC: Independent Colleges Office.

Project Kaleidoscope. (1996). Structures for science: A handbook on planning facilities for undergraduate natural science communities. Washington, DC: Independent Colleges Office.

Reese, C. M., Miller, K. F., Mazzeo, J., & Dossey, J. A. (1997). National assessment of educational progress 1996 mathematics report card for the nation and the states. Washington, DC: National Center for Education Statistics.

Rotberg, I. C. (1998). Interpretation of international test score comparisons. Science, 280, 1030–1031.

Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (1997). A splintered vision: An investigation of U.S. science and mathematics education. Boston: Kluwer.

Schlechty, P. C., & Vance, V. S. (1983). Recruitment, selection, retention: The shape of the teaching office. Elementary School Journal, 83, 469–487.

Seldin, P. (1989, March). How colleges evaluate professors, 1988 vs. 1983. AAHE Bulletin, 41(7), 3–7.

Seymour, E., & Hewitt, N. (1997). Talking about leaving: Why undergraduates leave the sciences. Boulder, CO: Westview.

Seymour, E., Daffinrud, S. M., Wiese, D. J., & Hunter, A. B. (2000). Creating a better mousetrap: On-line student assessment of their learning gains. Report to Exxon-Mobil Foundation; also Paper to the National Meetings of the American Chemical Society, San Francisco, March 2000.

Sims, C. (1992). What went wrong: Why programs failed. Science, 258(5085), 1185–1187.

Strenta, C., Elliott, R., Matier, M., Scott, J., & Adair, R. (1993). Choosing and leaving science in highly selective institutions: General factors and the question of gender. Report to the Alfred P. Sloan Foundation.

Tagomori, H. T., & Bishop, L. A. (1995). Student evaluation of teaching: Flaws in the instrument. Thought and Action: The HEA Higher Education Journal, 11(1), 63–78.

Tagomori, H. T., & Bishop, L. A. (1994). Content analysis of evaluation instruments used for student evaluation of classroom teaching performance in higher education. Paper to the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 1994.

Tapia, R. (1998). Assessing and evaluating the evaluation tools: The standardized test. Paper and presentation to the National Institute for Science Education Forum, Indicators of success in post-secondary SME&T education: Shapes of the future, February 23–24. In Millar, S. B. (Ed). 1998. Synthesis and Proceedings of the Third Annual NISE Forum, University of Wisconsin-Madison, NISE, WCER, pp. 13–21, 119–126.

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

Tobias, S. (1992). Revitalizing undergraduate science: Why some things work and most don't. Tucson, AZ: Research Corporation.

Treisman, U. (1992). Studying students studying calculus: A look at the lives of minority mathematics students in college. The College Mathematics Journal, 23(5), 362–372.

Trout, P. A. (1997). What the numbers mean: Providing a context for numerical student evaluation of courses. Change, 29(5), 24–30.

Turner, J. H. (1982). The structure of sociological theory, (3rd ed.). Homewood, IL: The Dorsey Press.

Van Dusen, G. C. (1997). The virtual university: Technology and reform in higher education. ASHE–ERIC Higher Education Report (Vol. 25, No. 5). Washington, DC: George Washington University.

Weise, D. J., Pedersen-Gallegos, L., Seymour, E., & Hunter, A. B. (2001). The role of the teaching assistant in undergraduate chemistry reform. Paper to the 6th Gordon Conference on Innovations in College Chemistry Teaching, Ventura, CA, Jan. 2001.

Wiese, D. J., Seymour, E., & Hunter, A. B. (1999). Report of a panel testing of the student assessment of their learning gains instrument by faculty using modular methods to teach undergraduate chemistry. Report to the Exxon-Mobil Educational Foundation, May.

Weiss, I. R., Matti, M. C., & Smith, P. S. (1994). Teacher education in transition: Alternate certification, Texas-style. (Available in ERIC.)

Williams, W. M., & Ceci, S. J. (1997). How am I doing? Problems with student ratings of instructors and courses. Change, 29(5), 12–23.

Wilson, J. M. (1994). The CUPLE physics studio. Physics Teacher, 32(9), 518–523.

Wilson, M., & Mosher, D. N. (1994). Interactive multimedia distance learning (IMDL2 The prototype of the virtual classroom. In Proceedings of ED-MEDIA, 94—World Conference on Educational multimedia and Hypermedia, Vancouver, BC, June.