Linking Evidence and Promising Practices in Science, Technology, Engineering, and Mathematics (STEM) Undergraduate Education

A Status Report for

The National Academies National Research Council Board of Science Education

James Fairweather
Center for Higher and Adult Education
Higher, Adult and Lifelong Education
Michigan State University
Fairwea4@msu.edu

1. Introduction

During the past 20 years or so the National Science Foundation (NSF), the National Academies of Science (NAS), and professional societies such as the Accreditation Board for Engineering and Technology (ABET) have expanded their policy focus beyond traditional support for basic and applied research in Science, Technology, Engineering, and Mathematics (STEM) to include improving the quality of undergraduate teaching and student learning in these disciplines. This expanded mission is in part a response to the decline in students choosing to major in STEM fields, declining percentages of STEM undergraduates continuing to graduate school, and the social and economic consequences of these trends (National Science Foundation 1996; Center for Science Mathematics and Engineering Education, Committee on Undergraduate Science Education 1999). The recent workshops sponsored by the National Academies National Research Council Board of Science Education are meant to help redress this crisis in STEM undergraduate education.

As Seymour and Hewett (1997) have shown, poor teaching practices in college STEM courses appear to lie at the heart of some of these problematic trends. Not surprisingly the primary reform effort in undergraduate STEM education, whether funded by the external agencies such as the NSF or by individual institutions, has been at the classroom level particularly through the use of more effective pedagogical practices and a conceptual shift from teacher- to learner-centeredness. Most of the papers presented at the two NAS workshops focus on teaching and learning in the classroom as have most of the NSF-funded projects through its course and curriculum development (CCLI) program. Most university efforts to improve teaching and learning likewise focus on (voluntary) professional development programs for individual faculty members (Wulff & Austin, 2004).

Yet this broad agreement on “the problem” and the trend to focus on improving individual classroom instruction has not led to a more seasoned understanding of the root causes of ineffective teaching and learning, low retention in the major and the declining pool of American students entering into STEM doctoral programs. Most efforts to reform undergraduate STEM education start from a presumptive reform model, one based primarily on in-classroom innovation and the teaching-learning process. The premise here is that the collection of hundreds if not thousands of individual faculty member improvements, initiated at least in part by empirical evidence of effectiveness, will lead to an aggregate change of a high order of magnitude.

This view of educational reform rests on three additional premises, all related to faculty involvement in reform. First, empirical evidence of successful student learning outcomes, however measured, is a precondition for STEM faculty involvement in pedagogical reform. Second, pedagogical effectiveness varies by academic discipline thereby requiring that many if not most instructional reforms be tested for their fit with distinct STEM disciplines. Third, the instructional role can be addressed independently from other aspects of the faculty position, particularly research, and from the larger institutional context.

We have every reason to be wary of these premises. NSF- and association-funded reforms at the classroom level, however well intentioned, have not led to the hoped for magnitude of change in student learning, retention in the major, and the like *in spite of empirical evidence of effectiveness* (Eiseman & Fairweather, 1996; Fairweather & Beach, 2002; Wankat, 2002). An evaluation of more than 400 NSF CCLI projects twelve years ago found significant empirical evidence of the educational benefits pedagogical reforms meant to foster student engagement in STEM courses, much less evidence of the dissemination of successful practices beyond the individual faculty member and even less of the spread of these proven innovations to other institutions (Eiseman & Fairweather, 1996). As a typical, more recent example, Fisher, Zeligman, and Fairweather (2005) found that pedagogical reforms in engineering service courses dramatically improved ABET-derived student learning outcomes, including problem-solving and analysis of complex problems. Despite empirical evidence of effectiveness, the reforms died because no other faculty members were willing to invest the time to teach the course in the new manner in part because the time commitment was greater than for traditional lectures.

Periodically the NSF comes up with more systemic efforts at reform, such as with programs to establish post-secondary centers (e.g., the Center for the Integration of Research, Teaching and Learning) and engineering coalitions (e.g., ECSEL). These efforts tend to have transitory effects with benefits accruing to participating institutions but few beyond them (Eiseman & Fairweather, 1996; Fisher, Fairweather, & Amey, 2003).

Moreover, although faculty in STEM disciplines vary substantially on a broad array of attitudinal and behavioral measures (Fairweather & Paulson, 2008) careful reviews of the substantial literature on college teaching and learning suggest that the pedagogical strategies most effective in enhancing student learning outcomes *are not discipline dependent* (Pascarllea & Terenzini, 2005). Instead, active and collaborative instruction coupled with various means to encourage student engagement invariably lead to better student learning outcomes irrespective of academic discipline (Kuh et al., 2005, 2007). The assumption that pedagogical effectiveness is disciplinary-specific can result in “reinventing the wheel,” proving yet again that pedagogies engaging students lead to better learning outcomes.

Finally, resistance to adopting more effective teaching strategies in part derives from the perception of STEM faculty that the teaching process is at odds with the research process, and that research is more interesting and more valued at their institutions (Fairweather 1996; Massy, Wilger, & Colbeck 1994). The perception of the importance of teaching in faculty rewards and the perceived consequence of spending more time on improving teaching, namely having less time for research, adversely affects faculty involvement in pedagogical reform (Fairweather 2005). This behavioral pattern holds true *even when faculty members express a deep commitment to teaching and to their students* (Leslie, 2002).

In this paper I first examine the variety of problems in STEM education and the role of pedagogical reforms in solving them. Next I describe three ways to think about pedagogical reform in STEM. In addition I discuss in more detail the premise on which the NAS initiative for college STEM pedagogical reform rests, namely the role of empirical evidence in fostering improvement in STEM undergraduate education. I follow this section with a review of what we know about effective STEM instructional practices relying especially on the papers presented at the NAS workshops, including a discussion of the nature of the evidence presented about educational impact. I conclude with a discussion of the influence of the multifaceted nature of faculty work on pedagogical reforms as well as recommendations for future interventions.

2. STEM Educational Reform: Solutions in Search of a Problem

Drawing on the work by Seymour and Hewitt (1997), much of the research agenda sponsored by the NAS and NSF focuses on reforming STEM instruction in the college classroom. Yet improved classroom instruction addresses only part of the laundry list of problems in STEM education.

- *Increasing public awareness about STEM:* Attaining this goal requires expanding the role of STEM faculty to include outreach beyond the university. Achieving this goal may require STEM faculty to change the way they relate to nonscientists outside the university; it does not require changing instructional practices in the college classroom.
- *Increasing the STEM pipeline:* The STEM pipeline runs from K-12 through can doctoral programs and professional practice. Increasing the interest of K-12 students in the sciences rests on university outreach efforts and improved K-12 instruction, not on college teaching. In contrast, attracting students to STEM majors and keeping them in the major can be affected by the classroom experience (MacDonald & Korinek, 1995).
- *Enhancing the preparation of STEM college students for their professions:* Although there is some evidence that classroom experiences can enhance knowledge of the major and careers, preparation for the profession seems more a function of the curriculum and of co-curricular experiences such as co-op in engineering (Fairweather et al., 1996; Moore et al., 2000).
- *Improving student learning in STEM:* Improving teaching and learning in STEM classrooms is most clearly aligned with this goal. Even so, many of the more

complex learning objectives, such as the retention of knowledge over time, the application of knowledge to solve unfamiliar problems, and commitment to lifelong learning must be assessed in subsequent learning experiences rather than immediate classroom environment.

Making explicit the nature of the problem addressed is crucial to assess adequately the success or failure of any STEM classroom or curricular reform effort. In particular, when the problem requires solutions beyond improved classroom instruction, the reform effort, even when successful in improving student learning, may fail to resolve the larger problem (Fisher, Fairweather, & Amey, 2003).

3. Improving Student Learning in STEM

Reform efforts to improve STEM college classroom instruction typically address one of three distinct objectives, each with different implications for faculty and students.

- *Improving teaching:* Much research on improving STEM teaching is based on a teacher-centered concept of teaching and learning. Here the goal is first to identify benchmarks derived from the literature on effective instructional practices (e.g., Weimer, 1996). Improvement is judged by the degree of fit between what instructors do in the classroom with instructional approaches found in the literature to improve student learning rather than on a direct assessment of student learning outcomes.
- *Improving student learning:* Adopting what Barr and Tagg (1995) call a "learner-centered paradigm," instructional reforms focus on determining increases in student learning resulting from instructional changes. The gold standard here is the extent of improved student learning outcomes rather than change in

instructional procedures (although the two clearly are linked). As discussed in the previous section, “student learning” is not a singular goal (e.g., content knowledge, synthesis, problem-solving) nor is it necessarily limited to a single classroom setting. Studies focused on improving student learning tend to take an “optimization” approach, i.e., where maximization of a learning outcome is the standard by which the reform is measured irrespective of cost and implementation difficulties.

- *Improving student learning productivity:* D. Bruce Johnstone (Johnstone & Mahoney, 1998) coined the phrase “student learning productivity” to focus on aggregate student learning relative to cost. Here the goal is to identify the least expensive and easiest way to attain the greatest increase in student learning across a designated group (program, department, introductory courses). The focus is less on optimizing performance in a given classroom and more on the greatest gain for a given investment. The implication of this perspective for assessing student outcomes is profound. It may be that the greatest gain in aggregate student learning in STEM is achieved not through the adoption of optimal teaching practices in each classroom but through the elimination of the worst practices. Encouraging the majority of STEM faculty members who only lecture to use *any* form of active/collaborative instruction (not necessarily the optimal forms of these instructional approaches) may lead to much greater gains in student learning productivity.

The success of any STEM classroom reform rests in part on which of these outcomes we measure. The first and second objectives, respectively, focus on how well faculty members improve their teaching relative to optimal instructional standards or how much student learning improved. Most of the papers presented at the NAS workshops center on these two objectives, which emphasize the identification of best or optimal instructional practices in STEM. In contrast, learning productivity focuses more on getting STEM faculty members to stop using ineffective pedagogies and to adopt the most cost-effective of the known effective instructional alternatives. Learning productivity is best examined by considering the *aggregate* performance of a STEM academic program or department rather than the performance of individual faculty members.

Most faculty professional development programs use effective instructional strategies as benchmarks to guide changes in teaching practices (Colbeck, O'Meara, & Austin, 2008). Many of the faculty members drawn to these professional development programs are good teachers seeking to make their classroom instruction even better (Gappa, Austin, & Trice, 2007). Think of them as seeking to improve from an 80% to a 95% level of instructional efficiency. In this context we may see some modest improvement in STEM education at the classroom level because of greater use of effective instructional practices.

These professional development strategies, however, have two flaws when it comes to large scale improvement in STEM education. First, faculty members participating in these professional development activities seem to be committed to improving their teaching and likely already use somewhat successful pedagogies. These faculty members, many of whom presented papers at the NAS workshops, as a rule *are not causing the poor instructional outcomes in STEM.* As such we can expect only a modest improvement in student learning overall from strategies seeking to promote use of "best" instructional practices. Second, the majority of STEM faculty, which as a collective make the least use of the active and collaborative instructional methods widely found to increase student learning (Fairweather & Paulsen, 2008), are not seeking to optimize their instructional practice. Rather they seek to "satisfice" teaching in order to optimize the time they have for research (Massy & Zemsky, 1994). The greatest gains in learning productivity are likely to come from finding ways to engage this large group of STEM faculty in *any* form of pedagogy that increases student engagement. Such strategies depend less on identifying optimal pedagogies through additional classroom research and more on enhancing the value of teaching in faculty rewards (Fairweather, 2005) and on finding ways to get STEM faculty members to use instructional methods *already known to improve student learning*.

4. The Role of Evidence in STEM Educational Reform

Most of the research on STEM education presumes that empirical evidence is crucial to the adoption of effective reforms. This premise rests on the belief that STEM faculty members will not change their instructional approaches without evidence that alternatives are more effective. From this perspective we must emphasize the adequacy of research design and evaluation methods in documenting effectiveness. Ebert-May et al.'s paper, for example, shows that the effectiveness of workshops is a function in part of the methods used to assess their impact. Gijbels et al. show that how outliers (research studies) are treated in a meta-analysis affects the conclusions about the effectiveness of instructional innovations.

Yet research shows that acceptable research evidence of instructional effectiveness is a necessary but not sufficient condition for the adoption of educational reforms in STEM. Dancy & Henderson discuss the barriers to dissemination of STEM educational reforms go far beyond an individual teacher in a classroom. The ECSEL engineering coalition and the Center for the Integration of Research, Teaching and Learning (CIRTL) show that many other factors affect the adoption of instructional innovations. These include faculty work load, faculty rewards, sequence of courses in curricula, leadership, and resources (Fairweather, 1996). An evaluation of the NSF CCLI program (Eiseman & Fairweather, 1996) and reforms in engineering service courses (Fisher, Fairweather, & Amey, 2003) found similar results.

The question remains, how important is evidence of instructional effectiveness in reforming STEM undergraduate education? In this section I examine three assumptions on which the premise of the importance of additional empirical research evidence in reforming STEM teaching and learning rests.

- *Do we need more evidence about the effectiveness of active and collaborative teaching strategies and related efforts to foster student engagement in their own learning?* Here the answer is a definitive *no*. The general literature on college teaching and learning as summarized by many authors, Pascarella & Terenzini (2005) and Kuh et al. (2005) prominently among them, provides clear research evidence that active and collaborative instructional strategies are more effective than traditional lecture and discussion across most if not all dimensions of student learning.

- *Do we need more evidence about the relative effectiveness of particular types of active and collaborative instructional strategies?* Perhaps. If the reform goal is to help STEM faculty members already committed to effective instruction to choose the better of two pedagogical options then evidence about their relative effectiveness may be useful. As I stated above, these faculty members by and large are not causing the problems in STEM education. In contrast, we do not need additional evidence that almost any active or collaborative approach will result in better learning outcomes than the dominant pedagogical approach in STEM, the lecture. Such evidence already exists. As Massy & Zemsky (1994), David Leslie (2002), and I (Fairweather, 2005) have shown, faculty members currently unwilling to engage in newer, more effective pedagogical practices are unlikely to change their instructional approach because of empirical evidence. Instead, these faculty members respond to the larger reward structure in which they work. The key levers to promote changes in attitudes and behavior toward teaching among this large group of STEM faculty is more likely to rest on work allocation and rewards than on evidence of instructional effectiveness. After all, existing evidence about the relative ineffectiveness of the dominant teaching method in STEM, lecturing, has not led to dramatic changes in the use of that technique.
- *Do we need to replicate research on effective instructional practices from other academic disciplines in STEM?* Perhaps. Much evidence already exists but not in easily accessible formats. Some disciplines do not have teaching-oriented journals. Even in disciplines that have such journals, research-oriented faculty

members are unlikely to read them. In addition, much of the literature on teaching effectiveness lies in “gray literature” such as evaluation reports to the NSF, which has a limited distribution. Any replication of effective teaching in STEM is likely to be focused on demonstrating yet again that these techniques work and more on getting STEM faculty engaged in their teaching. It is unlikely, however, that this additional research will be more useful in promoting faculty engagement in teaching than say, changes to resources and work allocation.

In summary, although additional research may to useful to fine tune applications in specific STEM instructional settings and sub-disciplines, much research on effective teaching and learning already exists. The key to improving STEM undergraduate education lies in getting the majority of STEM faculty members to use more effective pedagogical techniques than is now the norm in the these disciplines. Additional research evidence will play only a small role in this process.

5. What We Know and STEM Teaching and Learning: Lessons for Future Research and Practice

As mentioned previously, there is a substantial literature on effective college teaching and learning (e.g., Pascarella & Terenzini, 2005). Most of this work centers on student engagement either in active and collaborative instruction in the classroom (e.g., Kuh et al., 2005) or in out-of-class learning environments (Brower & Inkelas, 2007). In addition, there is a large literature on faculty professional development (e.g., Wulff & Austin, 2004). In this section I focus on the contributions made by the NAS workshop papers and the implications for future research and practice.

The NAS papers are representative of research in STEM undergraduate education. Most focus on classroom teaching, one or two on curriculum reform, and one or two on professional development. Two papers direct attention to the instructional reform process at both the departmental and institutional levels. Only one paper examines the effect of out-of-class experiences, such as undergraduate research programs and co-op experiences in engineering, on STEM student learning.

The forms of assessment also represent the range typically found in the literature. Some papers describe courses and evaluation strategies to assess effectiveness but do not provide evaluative data. Others provide descriptive information about impacts but little else. Other authors use meta-analysis to examine the literature for evidence of effects across projects. A few focus on the psychometrics of instrumentation and the consequences for evaluation. Most estimates of effects on student learning in the papers are correlational, not experimental. A very few use classic pre-post tests to assess gains in student learning. One or two employ longitudinal or at least long-term cross-sectional studies to examine effects.

Whether in the form of problem-based learning (PBL), case-based learning, ways to build learning communities, or ways to shape professional development programs, the NAS workshop papers for the most propose active/collaborative instructional strategies for which we already have much empirical knowledge (e.g., Pascarella & Terenzini, 2005; Kuh et al., 2005, 2007). Most do not break new ground here, instead intending to translate more general effective practices into STEM applications. In doing so, some authors seem aware of the broad literature of teaching effectiveness and recognize the link between their research and the broader literature. Others seem less aware of relevant work done elsewhere.

Below I list highlights and lessons learned from the papers to inform future work in STEM undergraduate education.

- As a field of study, STEM education would benefit from distinguishing between educational topics where we have sufficient empirical evidence of success to warrant implementation without additional research, such as PBL, from those where additional research is needed.
- In addition to evidence of effectiveness, to encourage other STEM faculty members to adopt a particular instructional approach research should provide sufficient detail about the course, setting, resources required, and the like for others to see the relevance of the approach to their own work.
- A description of the strategies to implement reforms, whether in a classroom or curriculum, are as crucial to convincing potential adopters to try the new approaches as evidence of effectiveness.*If the goal is to encourage widespread use of effective instructional techniques then a description of what it takes to implement the innovations is as crucial as the evidence of their effectiveness.*
- Evaluation (and research) on instructional innovations should distinguish between the requirements for effective teaching techniques (e.g., providing clear learning objectives for students is important in all types of instruction) from requirement for more innovative forms of instruction (e.g., studio-based learning).
- Many of the classroom research projects receive substantial external funding to develop materials and assess outcomes. Most potential adopters will not have

similar resources. In this context researchers need to demonstrate the possibility of achieving similar effects *without* having e.g., externally-funded graduate assistants and thousands of dollars to develop materials. At the very least they should delineate the potential costs of implementing the reform to enable potential adopters to obtain the necessary resources before proceeding. Finally, materials, assessment instruments, and so on must be sufficiently *efficient* to make it possible for others to use them in their own classrooms. For example, efficiency and ease of are among the more important reasons for the widespread use of the FCI in physics.

- The usefulness of any assessment technique ultimately depends on both its rigor *and* ease of use. Labor intensive assessment tools requiring substantial training prior to use are not likely to find widespread acceptance among potential faculty adopters. Rather a more routine and easy to apply technique is crucial to widespread use. Researchers and faculty members who customize evaluation tools for idiosyncratic applications are not likely to find an enthusiastic response from colleagues. The FCI, mentioned above, is a good example of a research-based instrument whose widespread acceptance is in large measure a function of its ease of use.
- Longitudinal studies of effects are rarely carried out. Such studies, however, provide the ultimate basis for determining the effectiveness of changes in student learning (Connolly, 2008).
- Measuring student outcomes with self-assessed instruments, particularly standard student ratings of the classroom, runs the risk of showing negative effects when

students experience active and collaborative instruction for the first time. In retrospect the same students often increase their rating when they see the long-term effect of the experience on their ability to apply and retain knowledge (Fairweather et al., 1996).

- Student portfolios are potentially useful indicators of student progress. The key is showing potential adopters of educational innovations how to evaluate them.
- The most valid forms of assessing the instructional impact on learning break the student outcomes into components—e.g., knowledge content, problem solving, communication skills—and study the change in these outcomes separately rather than presenting a single measure of change in student learning. Each distinct outcome may (or may not) require a different assessment tool.
- The Gillian-Daniel paper is an excellent example of delineating the type of student impact and relating it to type of intervention and evidence. The likely impact of this type of paper is enhanced when the author includes the types and elements of data collected.
- Researchers should develop assessment procedures to distinguish between the types of learning outcomes meaningfully measured in a single class—e.g., content knowledge—from the types of learning outcomes (e.g., application of knowledge) best assessed in subsequent courses.
- Course and curriculum innovators should understand that there is no agreed upon “standard” for judging impact or statistical relationship between predictors and outcomes *among the community of potential STEM faculty adopters*. Instead, the field of STEM educational research very much needs to develop a consensus

about the important of research results. For example, using PBL may result in a small, significant change in pre-post test results of student learning at $p < .05$. Yet the *amount* of improvement may be too small to convince others to try PBL even though the results are statistically significant. The key here is the *meaningfulness* of the magnitude of change in the eyes of potential adopters, not the level of statistical significance important to referees of journal articles.

- Meta-analysis, which conceptually useful, can be problematic because of the wide variety of settings and sample sizes in research on teaching and learning. Gijbels et al. show that the decision how to handle outliers can change the effect size from one showing benefit to one showing little change. Ebert-May et al. show that the type of evaluation (e.g., self-reporting versus observation) can affect conclusions about effectiveness. A crucial task in any STEM educational reform effort is to ensure that the findings about effectiveness are *not* artifacts of the evaluation technique.
- Effective evaluation of STEM educational innovations requires determination of when student and faculty attitudes and perceptions are relevant measures of outcomes and when more traditional measures, such as test scores, are required. Feldman (1998), as one example, is a useful guide for judging the adequacy of student ratings of classroom teaching. As another example, Lattuca and Terenzini (see Volkwein et al., 2004) have studied the reliability and validity of different ways of measuring changes in student outcomes based on ABET EC2000 criteria.
- The tradeoff in conducting in-depth studies of single settings, whether a course, department, or institution, is between the rich detail available when studying one

location, detail without which it may be impossible to detect effects, and the potential limited generalizability of results to other locations. One way to address generalizability here is to include sufficient detail about the site to help others assess its relevance to their circumstances.

- Much of the research on STEM education assumes the greatest variation in results is a function of discipline, e.g., engineering versus chemistry. This assumption may be false (Fairweather & Paulson, 2008). It may be that the instructional setting—laboratory, lecture hall, recitation—is a more important influence on student learning outcomes than academic discipline. Similarly, the effectiveness of instructional practices may vary considerably between courses for majors and non-majors (Fisher, Zeligman, & Fairweather, 2005).
- Many of the workshop papers demonstrate knowledge of the teaching literature in their sub-disciplines but less so about the general body of knowledge about teaching and learning. This point is consistent with Darcy & Henderson's literature review.
- Few papers provide sufficient formative assessment or a description of implementation steps to help potential adopters put the innovative programs or practices in place.
- Fit the type of evaluation data to the audience. For some audiences student self-reports and surveys of faculty attitudes are sufficient. For others pre-post test data are the minimally acceptable standard for measuring effects.
- The role of face validity in the assessment of STEM education is not well understood. Stories about educational experiences do not meet the same scientific

standard for evidence as pre-post tests but they may be essential to helping potential adopters to see the fit between the innovation and her/his classroom. Alternatively, studies where stories are the only form of evidence are not likely to meet the evidence standard many STEM faculty members would accept.

- To be effective evaluations should distinguish between the type and format of information needed to help a faculty member (a) implement an innovation (formative assessment), (b) summarize the effectiveness of changes in teaching practices on student learning, (c) demonstrate effectiveness as a research project where the audience consists of national disciplinary peers, and (d) present findings to encourage other faculty members to change the way they teach.
- Curriculum reform requires more than evidence of instructional effectiveness. Relevant evidence must include cost-effectiveness and be sensitive to the politics of the academic program and department (Fisher, Fairweather, & Amey, 2003).
- Micro-level assessments, such as case-based instruction of student groups in studio settings, provide important information about the effectiveness of the immediate environment. They often do not address the importance of the larger environment, such as the support for the department chair, resources provided by the dean, and so on.
- To increase the implementation of instructional strategies shown effective requires a model of change, including the roles of research evidence, leadership, resources, faculty work load and rewards, and resources. In this context, empirical evidence is only one part of the reform effort. It is a necessary but not sufficient condition for improving teaching and learning in STEM.

- Sometimes a proposed instructional innovation may stimulate innovation without empirical evidence of effectiveness if the instructional problem is seen as widespread.
- Start every STEM educational reform and evaluation *as if it will eventually be disseminated and scaled up*. This approach is much more likely to lead to greater use of the instructional innovation than an evaluation focused solely on idiosyncratic factors in the classroom (or curricular) environment.

6. Reforming STEM Education: Understanding the Dynamics of Faculty Work

Promoting the use of effective instructional strategies by STEM educators requires understanding how teaching fits the world of faculty work. Evidence of teaching effectiveness alone is not sufficient because teaching does not happen in a vacuum. Teaching takes place in a social system where faculty members attempt to balance teaching, research, and service in a manner consistent with both the dominant mythology about faculty productivity and reward systems that assign relative values to teaching and research.

The Relative Value of Teaching and Research

The NAS and NSF initiatives to promote evidence-based effective college teaching and learning in STEM owe allegiance in part to the broad movement started by Ernest Boyer in *Scholarship Reconsidered* (1990). Boyer argued for a renewed commitment to college teaching to counteract what he saw as the over-emphasis on traditional faculty scholarship, particularly through refereed publications. Boyer believed that recasting teaching as a form of scholarship—assessing student learning and disseminating the knowledge to others—might reclaim some stature for teaching on major college campuses. Boyer's concern about the decreased value placed on teaching, especially in major universities, was not misplaced. Since the 1980s publishing prowess has been the key to internal (Fairweather, 1996) and external (Winston, 1994) labor markets and to promotion and tenure ((Braxton, Luckey, & Helland, 2002).

The efforts to promote effective college teaching have borne fruit. We have seen increasing research to demonstrate the effectiveness of active and collaborative instructional practices in improving student learning (Bruffee, 1993; Wankat, 2002), including the working papers sponsored by this NAS workshop. Many academic institutions have established centers for teaching and learning to encourage the use of these practices by college teachers (Rice, Sorcinelli, & Austin, 2000; Seldin and associates, 1990). The NSF-funded Center for the Integration of Teaching, Learning and other initiatives are leading efforts to promote effective instructional practices.

Yet countervailing forces remain. We have witnessed increased use of part-time and adjunct faculty in all types of institutions to carry out a large part of the instructional workload, leaving tenure-track faculty members free to focus more heavily on traditional scholarship (Baldwin & Chronister, 2001; Finkelstein, Seal, & Schuster, 1998). Despite evidence of broader definitions of acceptable forms of scholarship, Braxton, Luckey, & Helland (2002) found that published research remains the coin of the realm for promotion and tenure decisions in most four-year institutions.

Efforts to enhance the value of teaching on college campuses inevitably confront the prevailing reward systems. Although it is possible to attract some faculty to "the cause" by appealing to their better nature, for the most part it is difficult to attain widespread participation in faculty development and other teaching-related endeavors without indications that such efforts are valued by the institution. At the very least, faculty members must believe that the additional time they spend on developing new instructional strategies will not be punished when it comes time for annual reviews, raises, and promotion and tenure.

Here the evidence is less promising. Despite extensive efforts by the NSF, the creation of teaching and learning centers on many college campuses, the development national organizations to promote faculty professional development, and research evidence about the effectiveness of active and collaborative instruction in fostering student learning, faculty rewards remain significantly slanted in favor of research and publishing. This pattern is quite visible in the predictors of faculty salaries in the years since Boyer wrote *Scholarship Reconsidered*. As I demonstrated in previous work (Fairweather, 2005), the more time that faculty spend in the classroom teaching the lower the average salary. *This pattern holds true for all types of 4-year institutions, including liberal arts colleges.* In the six years between 1992-3 and 1998-9, the most recent data available from the National Center for Education Statistics, hours spent in the classroom teaching became more significantly, negatively related with pay *in the most teaching-oriented institutions*. In contrast, career publications remain the strongest predictor of faculty pay *irrespective of type of institution*. An economic analysis of the estimated effect of an additional hour spent in the classroom and an additional career publication at the mean shows that it costs money to spend time teaching whereas publishing is invariably rewarded with higher pay. On the basis of this research I concluded:

> In sum, despite decade-long efforts to enhance the value of teaching in four-year colleges and universitiesthis study shows that spending more time on teaching, particularly classroom instruction, still means lower pay. Traditional scholarly productivity remains the strongest behavioral predictor of faculty pay although

> some trends suggest that its importance in pay may be stabilizing or even slightly decreasing… The declining monetary value of classroom instruction across types of institution should give us all pause to consider the fit between our rhetoric about the value of teaching and the rewards actually accrued by faculty who teach the most. Especially troubling is the declining value of classroom teaching over time in teaching-oriented institutions...patterns in faculty pay suggest that institutional decision-makers treat teaching and research as discrete categories of work when assigning salaries. Whether intentional or not, the aggregate consequences of these individual decisions about salaries seemingly set teaching in opposition to research when it comes to the monetary value that institutions place on them (Fairweather, 2005, pp. 420-421).

These findings strongly suggest that enhancing the value of teaching in STEM fields requires much more than empirical evidence of instructional effectiveness. It requires active intervention by academic leaders at the departmental, college, and institutional level. It requires efforts to encourage a culture within academic programs that values teaching. Whether through intervention in promotion and tenure decisions, salary structures, or provision of additional resources, active engagement by institutional leaders is a prerequisite to teaching reform efforts to succeed. In the end, faculty members take their cues about what their institutions value by looking at salary and promotion and tenure decisions rather than the rhetoric about or evidence in support of good teaching.

The Nexus between Teaching and Research

The dominant ideology in faculty work is that teaching and research are mutually reinforcing. This ideology is expressed primarily in the ideal of the *teacher-scholar* (Crimmel, 1984), which is institutionalized in the annual review process and in promotion and tenure decisions by expecting faculty members simultaneously to be productive in both teaching and research (along with some contribution to service):

> Teaching and research are seen as mutually reinforcing. From this perspective, the best scholars are the best teachers; the best teacher is a scholar who keeps abreast

> of the content and methods of a field through continuing involvement in research and who communicates knowledge and enthusiasm for a subject to students (Fairweather, 1996, p. 100).

Meta-analyses by Feldman (1987) and Hattie & Marsh (1996) found little evidence to support this ideology. Feldman found an average correlation of .12 between student ratings of teaching excellence and various forms of research productivity. Ten years later Hattie & Marsh found that the average correlation for this relationship had decreased to .06. Studies of time allocation consistently find a substantial negative correlation (r = -.50 or greater) between time spent on teaching and time spent on research (Fairweather, 2002). My study of faculty productivity (Fairweather, 2002) shows that the percentage of faculty members during a given two-year period simultaneously above the median on teaching (student credit hours generated) and research (publications in various forms) is 22.0%. This percentage does not vary substantially by institutional type.

These findings mean that more time and attention spent by STEM faculty pay to their teaching often comes at a cost to the time they have for research. Carefully crafted institutional policies are needed to integrate efforts to promote better teaching in this complex faculty work environment:

> Policies meant to encourage teaching productivity and effectiveness adversely affect individual research productivity, and vice versa. More complex and potentially successful policies might reward teaching and research productivity differently at distinct points in the faculty career. Alternatively, rather than having a single broad expectation for faculty work, academic policies might differentiate individual faculty responsibilities and allocate rewards accordingly (Fairweather, 2002, p. 44).

7. Intervention Strategies

On balance the research suggests that the greatest gains in STEM education are likely to come from the development of strategies to encourage faculty and administrators to implement proven instructional strategies rather than to carry out additional research on these strategies. Additional research may be helpful but it should be carefully targeted, perhaps under the guidance of the NAS and NSF, to fulfill gaps in existing knowledge rather than to prove yet again that active and collaborative strategies are more effective than passive ones.

The primary element of any implementation strategy must be to acknowledge that STEM educational reforms take place in a social context, one which typically rewards research more than teaching and asks faculty members simultaneously to be productive in research, teaching, and service. Appeals to STEM faculty on the basis of research evidence may well attract a small cadre of committed teachers yet most of these faculty members already are seeking ways to improve their teaching. The largest gain in learning productivity in STEM will come from convincing the large majority of STEM faculty that currently teaches by lecturing to use any form of active or collaborative instruction.

Among the most important elements of a successful change strategy to promote the improvement of undergraduate STEM education are the following:

- Develop distinct models for implementation, dissemination, and institutionalization for STEM reforms where the relative roles of evidence-based research on teaching, leadership, work loads, rewards, and so on are clearly delineated.

- Most professional development organizations are considered by faculty members as on the periphery of their academic work lives, part of the support services at colleges and universities. Regardless of their quality, these support units remain detached from faculty work lives unless department chairs and deans encourage their faculties to participate in them. The effective use of teaching and learning centers and the similar units depends in large part on the incorporation of these services into the mainstream of faculty work and rewards.
- External networks of like-minded colleagues outside of one's institution can be important forces in promoting instructional reform. The ECSEL Coalition and CIRTL made it possible for faculty members committed to STEM educational reform but isolated in their own work environments to find supportive colleagues elsewhere.
- Professional societies can be leveraged to enhance STEM education. The same is true for the NAS and NSF.
- Distinguish strategies aimed at reforming the instructional practice of current faculty members from strategies meant to inculcate a commitment to teaching by future faculty members, i.e., doctoral students.
- Identify key leverage points in the relevant change model. Efforts meant to assist STEM faculty members improve their teaching will have little effect if department chairs assign untrained TAs to teach the courses.
- Recognize that strategies aimed at "preaching to the choir" may not be effective for the average STEM faculty member. Faculty members deeply committed to reforming their teaching already make use of professional development programs,

seek research evidence to guide their choice of pedagogy, and so forth. The same strategies may not work for the faculty who are not so inclined.

- Above all, recognize that more effort needs to be expended on strategies to promote the adoption and implementation of STEM reforms rather than on assessing the outcomes of these reforms. Additional research can be useful but the problem in STEM education lies less in not knowing what works and more in getting people to use proven techniques.

References

Baldwin, R., & Chronister, J. (2001). *Teaching without tenure: Policies and practices for a new era.* Baltimore: Johns Hopkins University Press.

Barr, R., & Tagg, J. (1995). From teaching to learning: A new paradigm for undergraduate education. *Change 27*: 12-15.

Boyer, E. L. (1990). *Scholarship reconsidered: Priorities of the professoriate*. Princeton, N.J.: Carnegie Foundation for the Advancement of Teaching.

Braxton, J. , Luckey, W., & Holland, P. (2002). *Institutionalizing a broader view of scholarship through Boyer's four domains.* ASHE-ERIC Higher Education Report, Vol. 29, No. 2. San Francisco: Jossey-Bass/John Wiley Periodicals.

Brower, A., & Inkelas, K. (2007). Assessing learning community programs and partnerships. In *Learning communities and student affairs: Partnering for powerful learning*, ed. B. Smith & L. Williams. Washington, D.C.: National Association of Student Personnel Administrators.

Bruffee, K. (1993). *Collaborative learning: Higher education, interdependence, and the authority of knowledge.* Baltimore: Johns Hopkins University Press.

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

Colbeck, C., O’Meara, K, & Austin, A. (2008). *Educating integrated professionals: Theory and practice on preparation for the professoriate.* San Francisco: Jossey-Bass.

Connolly, M. (2008). Effects of a future-faculty professional development program on doctoral students and postdocs in science, technology, engineering, and math: Findings from a three-year longitudinal study. Paper presented at the Conference on preparing for Academic practice: Disciplinary perspectives, Oxford, England.

Crimmel, H. (1984). The myth of the teacher-scholar. *Liberal Education 70*: 183-198.

Eiseman, J., & Fairweather, J. (1996). *Evaluation of the National Science Foundation Undergraduate Course and Curriculum Development Program: Final Report.* Washington, D.C.: SRI International.

Fairweather, J. (1996). *Faculty work and public trust: Restoring the value of teaching and public service in American academic life*. Boston: Allyn & Bacon.

Fairweather, J. (2002). The mythologies of faculty productivity. *Journal of Higher Education 73*: 26-48.

Fairweather, J. (2005). Beyond the Rhetoric: Trends in the Relative Value of Teaching and Research in Faculty Salaries. *Journal of Higher Education 76*: 401-422.

Fairweather, J., & Beach, A. (2002). Variation in Faculty Work within Research Universities: Implications for State and Institutional Policy. *Review of Higher Education 26*: 97-115.

Fairweather, J., Colbeck, C., Paulson, K., Campbell, C., Bjorklund, S., & Malewski, E. (1996). *Engineering coalition of schools for excellence and leadership (ECSEL): Year 6*. University Park, PA: Center for the Study of Higher Education, Penn State University.

Fairweather, J., and Paulson, K. (2008). The Evolution of Scientific Fields in American Universities: Disciplinary Differences, Institutional Isomorphism. In *Cultural perspectives in higher education*, ed. J. Valimaa and O. Ylijoki (pp, 197-212). Dordrecht: Springer.

Feldman, K. (1987). Research productivity and scholarly accomplishment of college teachers as related to their instructional effectiveness: A review and explanation. *Research in Higher Education 26*: 227-298.

Feldman, K. (1998). Reflections on the study of effective college teaching and student ratings: One continuing quest and two unresolved issues. In *Higher education: Handbook of theory and research, Volume 13*, ed. J. Smart (pp. 35-74). New York: Agathon Press.

Finkelstein, M., Seal, R., & Schuster, J. (1998). *The new academic generation: A profession in transition.* Baltimore: Johns Hopkins University Press.

Fisher, D., Fairweather, J., & Amey. M. (2003). Systemic reform in undergraduate engineering education: The role of collective responsibility *International Journal of Engineering Education 19*: 768-776.

Fisher, P., Zeligman, D., & Fairweather, J. (2005). Self-assessed Student Learning Outcomes in an Engineering Service Course. *International Journal of Engineering Education. 21*: 446-456.

Gappa, J., Ausin, A., & Trice, A. (2007). *Rethinking faculty work: Higher education's strategic imperative.* San Francisco: Jossey-Bass.

Hattie, J., & Marsh, H. (1996). The relationship between research and teaching: A meta-analysis. *Review of Educational Research 66*: 507-542.

Johnstone, D., & Mahoney, P. (1998). Enhancing the productivity of learning: Curricular implications. *New Direction for Higher Education 26*: 23-34.

Kuh, G., Kinzie, J., Buckley, J., Bridges, B., & Kayek, J. (2007). *Piecing together the student success puzzle: Research, propositions, and recommendations.* Washington, D.C.: Association for the Study of Higher Education.

Kuh, G., Kinzie, J., Schuh, J., & Witt, E. (2005). *Student success in college: Creating conditions that matter.* Washington, D.C.: Association for the Study of Higher Education.

Leslie, D. (2002). Resolving the dispute: teaching is academe's core value. *Journal of Higher Education 73*: 49-73.

MacDonald, R., & Korinek, L. (1995). Cooperative learning activities in large entry-level geology courses. *Journal of Geological Education 43*: 341-345.

Massy, W., Wilger, A., & Colbeck, C. (1994). Department cultures and teaching quality: Overcoming "hallowed" collegiality. *Change 26*: 11-20.

Massy, W., & Zemsky, R. (1994). Faculty discretionary time: Departments and the "academic ratchet." *Journal of Higher Education 65*: 1-22.

Moore, K., Fairweather, J., Amey, M., Ortiz, A., Mabokela, R., & Ruterbusch, M. (2000). *Best practices for reform in undergraduate education in science, math, engineering, & technology: A knowledge framework.* East Lansing, MI: Center for the Study of Advanced Learning Systems, Michigan State University.

National Science Foundation. (1996). *Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology.* Washington, D.C.: National Science Foundation.

Pascarella, E., and Terenzini, P. (2005). *How college affects students: A third decade of research*. San Francisco: Jossey-Bass.

Rice, R., Sorcinelli, M., & Austin, A. (2000). *Heeding new voices: Academic careers for a new generation.* Washington, D.C: American Association for Higher Education.

Seldin, P., & associates (1990). *How administrators can improve teaching: Moving from talk to action in higher education*. San Francisco: Jossey-Bass.

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

Volkwein, J., Lattuca, L., Terenzini, P., Strauss, L., & Sukhbaatar, J. (2004). Engineering change: A study of the impact of EC2000. *International Journal of Engineering Education 20*: 318-328.

Wankat, P. (2002). *The effective, efficient professor: Teaching, scholarship, and service.* Boston: Allyn & Bacon.

Winston, G. (1994). The decline in undergraduate teaching: Moral failure or market pressure? *Change*, *26*, 8-15.

Wulff, D., & Austin, A. (2004). *Paths to the professoriate: Strategies for enriching the preparation of future faculty*. San Francisco: Jossey-Bass.