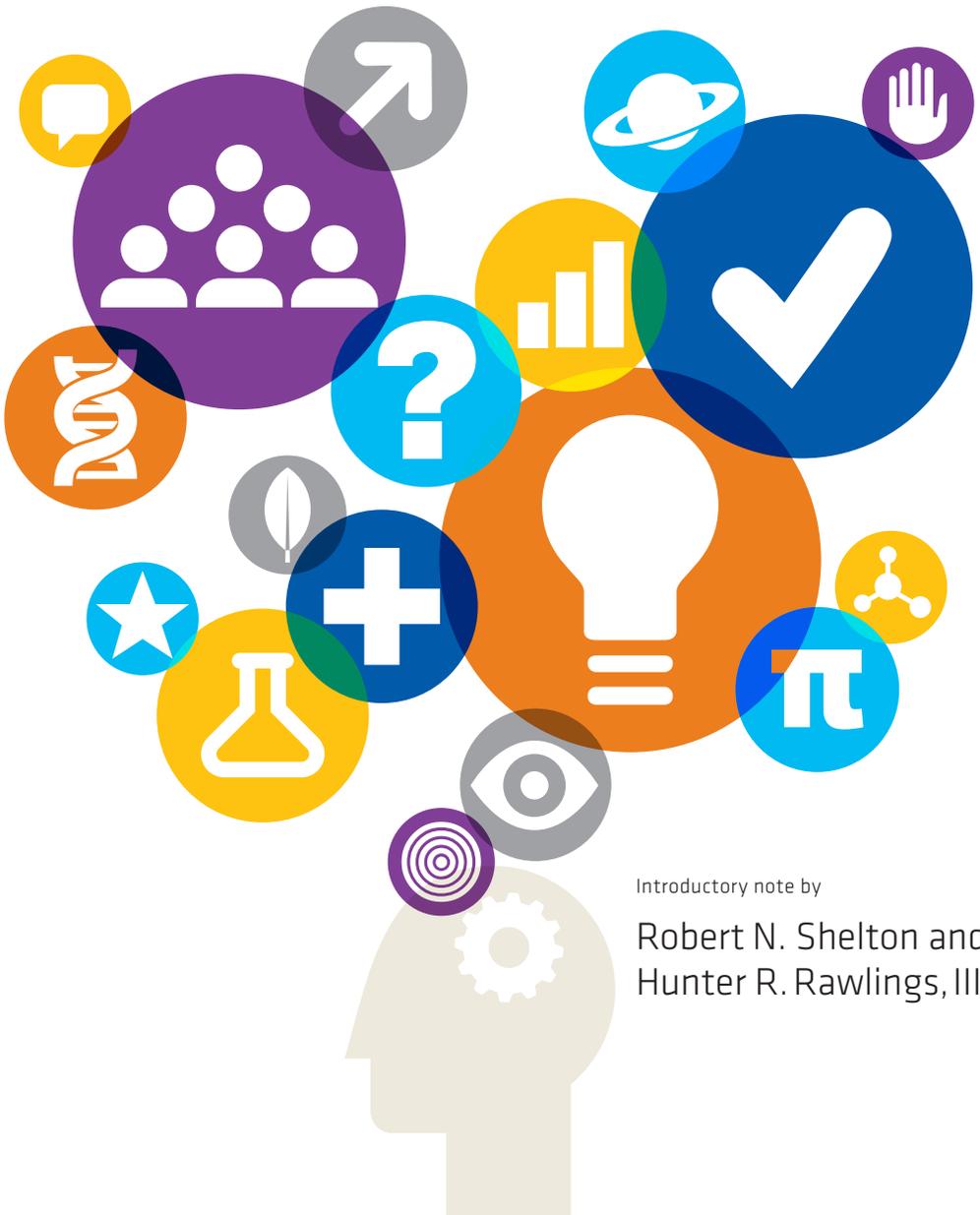




Published by Research Corporation
for Science Advancement

Searching for Better Approaches: Effective Evaluation of Teaching and Learning in STEM



Introductory note by

Robert N. Shelton and
Hunter R. Rawlings, III

Searching for Better Approaches: Effective Evaluation of Teaching and Learning in STEM

Introductory note by

Robert N. Shelton and
Hunter R. Rawlings, III



Published by Research Corporation
for Science Advancement

© 2015 Research Corporation
for Science Advancement
All rights reserved. First edition 2015.
*Not for quotation or reproduction without
permission of the authors or
Research Corporation for Science Advancement*

Chapter 9 © 2015 The IDEA Center

Design by Godat Design

ISBN 978-1-4951-6339-5


RESEARCH CORPORATION
for **SCIENCE ADVANCEMENT**
A foundation dedicated to science since 1912.

4703 East Camp Lowell Drive, Suite 201
Tucson Arizona 85712

Contents

Acknowledgements	vi
A Note from the Presidents	vii
Foreword	ix
Introduction	viii
Implementing Evidence-Based Undergraduate STEM Teaching Practices	1
Part I: Workshop Sessions	
1 Pre- and Post-Testing and Discipline-Based Outcomes <i>Adam K. Leibovich</i>	9
2 Student Identification of Learning Outcomes and Improved Student Evaluations <i>William R. Dichtel</i>	13
3 Peer Observation and Evidence of Learning <i>Andrew L. Feig</i>	21
4 Analytics and Longitudinal Assessment <i>Stephen E. Bradforth</i>	27
5 Administration and Implementation: Incentivizing, Uses, and Abuses of Evaluation and Assessment <i>Emily R. Miller</i>	33
Part II: Submitted Papers	
6 Pre- and Post-Testing in STEM Courses	39
7 Closing the Gap between Teaching and Assessment	43
8 Improving Teaching Quality through Peer Review of Teaching	47
9 Student Ratings of Instruction in Lower-Level Postsecondary STEM Classes	59
10 Enhancing Teaching and Learning: Potential and Reality	73
11 In Search of Improved Student Learning: Strategies for Affirming We Are “On Track”	81
Workshop Participants	87
About the Sponsors	88

Acknowledgments

The summary presented in this publication is the result of a two-day workshop, “Effective Evaluation of Teaching and Learning,” held in January 2014. The workshop, organized by the Cottrell Scholars Collaborative and the Association of American Universities (AAU) and supported by funds from the Research Corporation for Science Advancement, brought together leading research-active faculty as well as higher education scholars and practitioners to discuss how to more effectively evaluate teaching and learning in undergraduate STEM education. We are grateful for their insights and contributions to the workshop.

We also want to acknowledge the authors of commissioned papers and the members of the workshop planning and editorial team for playing a key role in the development of this publication: Stephen E. Bradforth, Department of Chemistry, University of Southern California; William R. Dichtel, Department of Chemistry and Chemical Biology, Cornell University; Adam K. Leibovich, Department of Physics and Astronomy, University of Pittsburgh; Andrew L. Feig, Department of Chemistry, Wayne State University; James D. Martin, Department of Chemistry, North Carolina State University; Karen S. Bjorkman, Department of Physics and Astronomy, University of Toledo; Zachary D. Schultz, Department of Chemistry and Biochemistry, University of Notre Dame; and Tobin L. Smith, Association of American Universities. Special thanks go to Emily R. Miller, Association of American Universities for her leadership in serving as the project coordinator for the workshop and for this publication.

Finally we are grateful to the Association of American Universities for hosting the workshop and for logistical support.

This publication is based upon work supported by the Research Corporation for Science Advancement; the views and recommendations expressed within this material are those of its authors and do not necessarily represent the views of the Research Corporation for Science Advancement.

A Note from the Presidents

In 2011, the Association of American Universities (AAU), an association of leading research universities, launched a major initiative to improve the quality of undergraduate teaching and learning in science, technology, engineering, and mathematics (STEM) fields at its member institutions. That same year, Research Corporation for Science Advancement announced the creation of the Cottrell Scholars Collaborative—a network of more than 250 outstanding teacher/scholars in the physical sciences who have received Research Corporation’s Cottrell Scholar Award—to work in teams and with other national initiatives on projects aimed at overcoming longstanding impediments to excellence in teaching and student learning at colleges and universities.

At the 2012 Cottrell Scholars Conference, AAU staff and a group of Cottrell Scholars focused on a major barrier to improving the quality of undergraduate education: the predominant use of student-based evaluations to assess teaching quality at colleges and universities. While effective at assessing faculty popularity, these student evaluations often fail to reflect accurately teaching quality and student learning. The two groups proposed a joint project, which Research Corporation supported, aimed at identifying new and innovative means to evaluate and reward teaching quality.

In January 2014, to help accomplish the goals of this project, AAU and a subset of scholars from the Cottrell Scholars Collaborative held a joint workshop on the effective evaluation of undergraduate teaching and learning. What follows is a summary of the results of that workshop, along with invited papers and reflections from workshop speakers and participants.

We hope that this report will help to inspire faculty members, whole departments, and entire institutions to use alternative methods for assessing quality teaching to supplement traditional student evaluations, and to seek out new and innovative ways to recognize and reward teaching excellence.

Sincerely,

Robert N. Shelton
President
Research Corporation for Science Advancement

Hunter R. Rawlings, III
President
Association of American Universities

Foreword: When Will Failure Be an Option: The Challenges of Teaching STEM in Colleges and Universities

Peter K. Dorhout,
Kansas State University

There are three numbers that you should keep in mind throughout this essay: 102, 276, and 5. Three numbers that tie together how we scientists might consider affecting change in how we teach and assess teaching Science, Technology, Engineering, and Mathematics, or STEM. Some add “Arts” to make this STEAM, and others add “Humanities” to make STHEM, or something like that. This essay will explore how, when one considers the methods of teaching at the turn of the 20th Century and the impact of engaging undergraduate students in the discovery process, there is much to be learned about how we teach science, how we can utilize feedback mechanisms to improve what we do, and how we can apply principles of leadership to change the status quo.

The first number: 102. It was roughly 102 years ago (1912) that Frederick Gardner Cottrell realized how difficult it was at the time to acquire funds to support fundamental research in science, and he created a unique philanthropic foundation, Research Corporation, to support basic scientific research.¹ Using funds that he garnered through a series of patents to manage the emissions from turn-of-the-century coal-fired plants—the electrostatic precipitator—a foundation was born. At a time when academic training and research was undergoing a renaissance of its own, and models of apprenticeships and research for undergraduates were under stress, academic scientists were finding it difficult to fund basic discovery research. Research Corporation was poised to fill a critical gap in educating the next generation of scientists.

So what of this period of discovery in primarily American science? There were a few institutions in the United States at the turn of the 20th Century, fewer still were public colleges and universities, where a student could learn and study science and engage in scientific research. The Land Grant legacy of public research universities had taken hold, still only about 40 years old, but the mechanisms for funding transformational research had been lagging the advances in university research. Private colleges and universities were ahead of the curve supporting research, but access to higher education was still very limited.

In 1925, Sinclair Lewis authored a book, *Arrowsmith*, that offered a glimpse into how science training and research was accomplished during

this transformational period in American education.² The fictional Martin Arrowsmith entered his university as a young man in 1904, focused on becoming a medical practitioner only to discover along the way how exploring fundamental science can lead to discoveries that chart a unique career path for those bold enough to follow.

Written about the period in history when Frederick Cottrell was making his first discoveries about electrostatic precipitation of fly ash and sulfuric acid mists, this story of Martin Arrowsmith highlighted a unique relationship between a student and a mentor, Dr. Max Gottlieb. Gottlieb was a physical chemist who believed that all things could be explained by science. Doubtless Gottlieb's optimism was fueled in part by Einstein's General Theory of Relativity, which was a mere nine years old at the time. For Arrowsmith and the reader, Gottlieb also represented the juxtaposition of science and religion at the time.

Arrowsmith was not a traditional protagonist in just any novel. Lewis wrote:

It cannot be said, in this biography of a young man who was in no degree a hero, who regarded himself as a seeker after truth yet who stumbled and slid back all his life and bogged himself in every obvious morass.

Isn't true science itself a belief system, a religion? Characterized by those things that we experience: stubbornness, desire, curiosity, sleeplessness, humility, do-your-best, failure, jubilation—if not religion, a very similar philosophy to religion. In the early 1900s, religion and science were often at odds, but they are paradoxically similar. Like religion, there is no national support for research, just benefactors. I will not cite the many philosophies of science promulgated by the likes of Plato, Aristotle, Bacon, and Kuhn, and which may be found in publications such as the journal *Philosophy of Science*, published by the University of Chicago Press since roughly this same time period. Nevertheless, there was a struggle of science and religion at this time that persists today in some ways.

Martin Arrowsmith endeavored to succeed under the mentorship of Max Gottlieb; he grappled with the fundamental understanding of science and the Scientific Method. He struggled so much that he penned his Prayer of the Scientist:

God give me a restlessness whereby I may neither sleep nor accept praise till my observed results equal my calculated results or in a pious glee I discover and assault my error.

Could it be that this universal prayer for restlessness of mind is something that we strive to impart to all our students who study science, whether or not we expect them to become scientists or to just appreciate science at some fundamental level? Could it be that this same restlessness is the state of mind that we educators of students seek—'restlessness whereby I may neither sleep nor accept praise until my students effectively learn the basic concepts from my instruction'?

So it has been over 102 years since Frederick Cottrell founded Research Corporation, a philanthropic foundation that supported 40 Nobel Laureates and thousands of fundamental studies in science that have impacted tens-of-thousands of students and faculty in physics, chemistry, and astronomy, to name just a few. In its eightieth year of incorporation, the Research Corporation board embarked on a renewal of sorts—its own renaissance of investment in research in higher education—and created the Cottrell Scholar program, a program to recognize and honor the philosophy of Frederick Cottrell.

The year 2014 marks the twentieth anniversary of this latest experiment in supporting research in higher education, which coincidentally began at the same time the National Science Foundation launched its CAREER program. Since its inception, the Cottrell Scholars program has recognized 276 Scholars—unique catalysts for change in STEM (or STEAM or STHM) education. **This is my second number to remember: 276.** These 276 Cottrell Scholars have mentored over 4,200 undergraduate and graduate students and taught thousands of course hours attended by tens-of-thousands of students. Cottrell Scholars are required to attend an annual meeting during the three years of their grant, meetings focused on:

- Collaborative connections to share best practices in teaching and research
- Leading new faculty workshops in chemistry and physics
- Focusing on how to teach science—grassroots efforts by junior faculty
- Mentoring junior faculty, postdocs, and students in more than 115 institutions

Let us return for a moment to the protagonist Martin Arrowsmith who struggled with the “control” experiment that led, ultimately, to failure for his phage treatment for the plague. His story ends after many personal and professional failures at finding a cure for the plague with Martin leaving the limelight of a research institute:

“I feel as if I were really beginning to work now,” said Martin. “This new [stuff] may prove pretty good. We’ll plug along on it for two or three years, and maybe we’ll get something permanent—and probably we’ll fail!”

So, from his vantage point as a young medical scientist, it didn’t matter if you failed, the truth was in the search and the belief that you may one day find something interesting. Arrowsmith discovered that if he didn’t become famous or have a great discovery, he was being true to himself. This is an important ethical lesson in science. It is the nature of our trade to fail and fail again—patience is the key. Failure, or learning how to fail and keep going, is something missing in our traditional classroom teaching environments; however, it is something that is key to our learning processes as scientists. We fail, and fail again, only to learn more about the true answers with each failure along the path to scientific enlightenment.

How do we teach patience? How do we teach science? How do we teach failure and persistence, and how do we assess learning when failure is part of the process? Is the ‘truth’ really in the search process, as Arrowsmith discovered? How do we teach the non-scientists that this is how science is done? How do we challenge the misconceptions—student contexts? How do we address cultural differences between students that lead to misconceptions?

Alas, this is not an essay with certainty in answers but one with more questions. Perhaps that makes it an essay on religion or philosophy. One thing is certain, in my opinion: the search for truth in how best to teach STEM is one argument for continuing the search. Assessing how students learn is part of the hypothesis and the feedback loop, and critical to improving our professions and disciplines. Therefore, we have an obligation to continue to proselytize for change in how we teach STEM.

It has been proposed that there are barriers for implementation of any new ideas or change in the current paradigm. Leading change in higher education relies on the five tenets or tent poles of leadership derived from the principles of the extraordinary leader put forward by J. H. Zenger and J. Folkman.³ **This is my last number for you to remember: 5.** According to Zenger and Folkman, these are the tenets of leadership displayed by successful leaders. In summary:

- **Character:** display high moral standards, demonstrate integrity
- **Personal Capabilities:** think proactively, evaluate progress
- **Focus on Results:** build consensus, utilize your skills, make work easy, apply evidence based practices, professionalize educational practice, ‘impedance match’ students with teaching
- **Interpersonal Skills:** listen before acting, team-building, learn how to use the ‘bully pulpit’, build support both up and down the leadership chain.
- **Plan for Success:** strategic thinking/planning, we are uniquely positioned to change the environment in higher education

As leaders in higher education, we are often called upon to do the unpopular, to change the status quo, and to test the truths about how students learn. If, as educators, we accept that we learn through many stimuli applied to us during our lives, and that our nature, culture, and local (and temporal) environment impact how we respond to stimuli, then it should be accepted that how we learn, as a person and as a species, changes with time. As scientists, should we not also accept that how we teach needs to change as our students and our subjects change? As science leaders in education, we should accept that we can apply the five tenets of leadership to our educational institutions to influence how STEM teaching can become part of a living pedagogy that relies on continuous assessment and input for regular renewal.

The only certainty I offer in conclusion is in the search for truth. The truth will elude us regardless how deeply we probe, teasing us to look deeper.

In a time when we read about scientific misconduct in the popular press and the stresses placed upon science when funding becomes more competitive, it is more important than ever to teach STEM through a combination of discovery, failure, new hypotheses, and deeper inquiries. ‘Getting our hands dirty’ or ‘observing the unexpected’ are proving harder to accomplish in an optimized college or university classroom setting, yet they remain sacrosanct in the practice and study of science. With generations of students who have received reward and recognition, been trained to avoid failure, and, to quote Garrison Keillor, “where all the women are strong, all the men are good-looking, and all the children are above average,” it will remain a challenge to teach students to embrace failure as a fundamental principle of science:

We’ll plug along on it for two or three years, and maybe we’ll get something permanent—and probably we’ll fail!”

¹ Research Corporation for Science Advancement, www.rcsa.org

² S. Lewis (1925) *Arrowsmith*, Harcourt Brace: NY

³ J. H. Zenger and J. Folkman (2002) *The Extraordinary Leader*, McGraw Hill: NY

Introduction

James D. Martin, North Carolina State University
Zachary D. Schultz, University of Notre Dame

Statement of Challenge

Teaching and learning are fundamental to the mission of higher education. It is incumbent on every professor and university to be vested in the success of that mission. At the intrinsic level our success is defined by the success of our students. At the national and global level effective teaching in science technology, engineering and math (STEM) disciplines is critical to ensure graduates are capable to address complex and interdisciplinary health, nutrition, energy and climate challenges.

In light of these challenges, the need for improving undergraduate education in STEM fields has received increased attention and taken on new urgency in recent years. The national policy environment reflects an increasing call to improve undergraduate teaching in STEM fields across and within relevant organizations.¹

The goal must be to improve, not just transform the current status of education. To accomplish this, effective tools for the evaluation of practice and personnel engaged in teaching and learning are critical to informed decision-making. Recent high-level policy reports have relied on effective teaching and learning research to identify deficiencies and potential solutions in STEM instructional practices and in institutional policies. Nevertheless, when it comes to evaluating teaching personnel, student teaching evaluations remain the de facto tool for assessment of teaching.

Current Practice of Evaluation of Teaching

An *Effective Evaluation of Teaching and Learning Survey* of chemistry, physics and astronomy department heads and faculty, discussed in Chapter 2, indicates student surveys are ubiquitously used for the evaluation of teaching in contrast to only a little more than half reporting the use of peer evaluation of teaching. Though implemented throughout the U.S. university system in the 1960s in response to student demands for a greater voice in their education, there is little evidence that student surveys have resulted in high quality teaching and learning.² Instead reliance on simplistic numerical student surveys often reinforces the status quo, makes faculty risk-averse with respect to experimenting with alternative pedagogy, and discourages faculty and administrators from thinking about or attempting reforms.

In our view, to achieve the goal of effective STEM teaching and learning the system has to change and the changes must be institutionalized. While some suggest that research universities and their faculty do not really

care about teaching standards,³ we believe that there is interest at both institutional and faculty levels to change, but the barriers to implementing such a culture shift are high. There is an increasing expectation that teaching practices be reformed, and should we fail to implement this cultural shift voluntarily, it is likely to be imposed upon us. State politicians and accreditation bodies will increasingly drive greater accountability measures and demand evidence that student learning is valued by faculty. When imposed our voices as STEM education practitioners will be a much more limited part of the conversation.

What is Effective for the Evaluation of Teaching and Learning?

So, what should be the scope for assessing teaching and learning and how should we do it in a valuable and informative manner? By employing a common research technique, we should consider a back-engineering thought process to develop strategies that will most effectively achieve the goal of effecting change to improve STEM undergraduate teaching and learning.

Backwards Engineering of Improved Teaching and Learning



Figure 1

Intermediate steps require definition and assessment of goals and strategies. Accomplishing this requires investment and commitment at all levels.

In defining learning goals, in our workshop, Robin Wright offered that effective teaching provides the student with the capacity to “use information after a significant period of disuse.” And Noah Finkelstein suggested the process of STEM education must be the “enculturation of students to become tool users.”

To understand where we are and where we are not meeting these goals, it should be clear that any evaluation structure must be expanded beyond simplistic student surveys. Measurements must assess both faculty teaching and student learning. And importantly, given that learning is not confined to a single course or the efforts of single faculty members, effective evaluation strategies must expand beyond assessment of the individual (teacher or student) to include systemic assessments of curricula and institutional structures.

With evaluations conducted, institutional cultures must change such that data does not get buried in reports that sit on shelves. Beyond doing evaluations, we must use the assessments to inform the development of capacity within faculty, students and the administration to affect change.

While effective change is generally developed and operationalized by practitioners (teachers and students), to ensure the improved changes are sustainable requires engaging institutional leaders such as department chairs, deans, and presidents in rethinking institutional structures and culture. The critical point of intersection is at the department level where cooperative practice must emerge, to create a culture that values thinking about and developing the best practice of teaching and learning.

Meeting the Challenge

To address these issues, a workshop sponsored by the Cottrell Scholars of the Research Corporation for Scientific Advancement (RCSA) and the Association of American Universities (AAU) was held in Washington, D.C. in January 2014. To ensure the “meeting in the middle” the workshop included Cottrell Scholars, teacher-researcher-scholar faculty from research-intensive institutions, along with specialists in science education and evidence-based learning, university administrators (ranging from provosts to department chairs), and disciplinary scientific society, association and federal research agency representatives, to explore how to best implement reforms to teaching evaluation practice at research intensive universities. This workshop focused on different methods individual schools have implemented to grapple with the challenges associated with creating a culture of positive STEM education. Together, faculty and administrators engaged to consider what does and does not work to identify, encourage, and reward effective teaching at their institutions.

Susan Singer, Director, NSF Division of Undergraduate Education, delivered the workshop’s keynote lecture and provided national context to a discussion of STEM teaching in higher education as well as the more general theme of implementing evidence-based teaching practices. Of these learner-centered practices, a piece that is often missing is the provision of data as feedback to a faculty member to calibrate how well they are teaching. This formed a significant component of the workshop’s focus. We concluded that any evaluation instruments and strategies must meet several criteria to garner the faculty respect necessary for broad adoption and to reform department cultures in a manner that elevates the value of effective teaching and learning within research-intensive universities. **In this report we attempt to capture the discussion and actionable practices that can be implemented across universities to assess student learning and improve instruction in STEM fields.**

Several example evaluation efforts that have proven successful were showcased; (1) having students complete pre- and post-tests as they move through an individual class and a course sequence; (2) tying faculty evaluation more closely to self-identified teaching goals and subsequent evidence of student learning; (3) department-wide implementation of peer observation of

teaching and faculty teaching portfolios; (4) effective strategies from university administrations and other external factors to incentivize departments to increase emphasis on effective teaching, and (5) implementation of longitudinal assessment and data analytics using course registration, student satisfaction and performance records to correlate student learning with the faculty teaching effectiveness.

At the heart of the workshop discussion was a central idea: “What would it look like if we evaluated teaching the way we evaluate research?” In this context, the workshop, and this report, are organized around a set of themes, from which we asked questions to frame the discussion and explore the successes, advantages, disadvantages and most important practicality for having real impact on student learning.

1: Pre- and Post-Testing and Discipline-Based Outcomes

- Can we address our teaching in a scholarly fashion without having to be discipline-based educational research (DBER) scholars?
- What do grassroots initiatives look like that apply our laboratory research and creative work to our teaching (and service)?
- How does one broadly implement known successful approaches to STEM education with non- or less-enthusiastic faculty?
- How can/should implementation of experimental discipline-based approaches to teaching be incorporated in teaching evaluations?
- Are there effective evaluation strategies that can be implemented broadly, given the realistic resources (financial and faculty time) available in most institutions?

2: Student Identification of Learning Outcomes and Improved Student Evaluations

- What are best practices for end-of-semester student surveys to maximize their value for evaluating student learning?
- What other techniques should be used for assessing student-learning outcomes, and what is a reasonable resource and time commitment needed to apply them within STEM departments in research universities?
- What are best practices that will allow formative use of the evaluation tools to help faculty improve their teaching?

3: Peer Observation and Evidence of Learning

- What are effective strategies to implement peer observation in a way that is not too demanding of faculty time?
- What training is required to ensure that peer observation actually assesses effective implementation of best pedagogical practices despite the fact that at the present, only a fraction of the faculty in typical STEM departments are currently champions of evidence-based teaching?

- What constitutes evidence for effective evaluation?
- How can the review be structured such that faculty are motivated to reflect on and assess their own teaching and implement change to continually improve?

4: Analytics and Longitudinal Assessment

- Are universities using the enormous amounts of data (student admission data, grade progression through college, class drop data, changes of major, time to graduation, and individual student's instructor evaluation history) effectively to improve teaching and learning?
- To correlate teaching with evidence of learning, or to identify choke points in specific academic programs, is the collection of new data needed? Or can this be accomplished with better use and broader access to data already collected?
- What data analytics techniques, such as cross-tabs and complex correlations, highly developed and used in business, are suitable to improve the evaluation of effective teaching and learning at the department, college and university levels?
- Can analytics provide accessible data in context that motivates faculty to prioritize improvement of their teaching and learning efforts?

5: Administration and Implementation: Incentivizing, Uses, and Abuses of Evaluation and Assessment

- What is the role of the department chair, dean, and ultimately provost in implementing change in assessment of student learning?
- What is the power of the bully pulpit to establish a culture of scholarly teaching and learning?
- How do the different levels of administration promote faculty buy-in?
- Is a faculty member's effort on effective teaching and learning appropriately recognized in promotion and tenure policy and practice?

Moving Forward

As faculty members, administrators and universities continue to engage with changes in the field of STEM education, we encourage our readers to think deeply about the practices used to evaluate teaching and learning. Change is hard and sometimes frustratingly slow to implement or to show results. To do it effectively, however requires a candid conversation. There will not be a one-size-fits-all solution to the evaluation of teaching and learning. Universities have their own constituencies and individual circumstances. They will be able to share practices and adopt those that work for them. Dissemination of these experiences is critical as the community strives to determine what combination of metrics provides sufficient granularity to understand what is happening in the classrooms and how that impacts student success without being so burdensome that effective implementation at scale is impossible.

Finally we note that any honest evaluation of the current status of teaching and learning is going to turn up some less than stellar policy and practice. Faculty members and administrators must be prepared to mutually challenge and provide the support and cover to actually wash the “dirty laundry” rather than hide it. As Provost Mary Ann Rankin reminded the workshop, we “must give patience and support as change happens.” While not everyone needs to be a discipline-based education research scholar, perhaps the first step to improve undergraduate STEM teaching is to expect all university faculty members to be scholarly about their teaching.

- 1 American Association for Advancement of Science (2011) *Vision and Change in Biology Education*. Washington, D.C.: American Association for Advancement of Science. <http://visionandchange.org/files/2013/11/aaas-VISchange-web1113.pdf>
S. R. Singer, N. R. Nielsen, H. A. Schweingruber (eds.) (2012) *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, D.C.: National Research Council. http://www.nap.edu/catalog.php?record_id=13362
National Science Board (2010) *Science and Engineering Indicators 2010*. Arlington, VA: National Science Foundation (NSB 10-01).
President’s Council of Advisors on Science and Technology. (2012) “Engage to Excel: Producing One Million Additional College Graduates With Degrees In Science, Technology, Engineering, And Mathematics.” http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_feb.pdf
- 2 P. B. Stark and R. Freishtat (2014) “An evaluation of course evaluations” *ScienceOpen Research* 2014 (DOI: 10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1) <https://www.scienceopen.com/document/vid/42e6aae5-246b-4900-8015-dc99b467b6e4;jsessionid=vwEToQjnLcXlWqpV2k9SeipM.master:so-app1-prd?0>.
M. Braga, M. Paccagnella, and M. Pellizzari (2014) Evaluating students’ evaluations of professors. *Economics of Education Review* 41, 71-88, doi: <http://dx.doi.org/10.1016/j.econedurev.2014.04.002>.
- 3 W. A. Anderson, U. Banerjee, C. L. Drennan, S. C. R. Elgin, I. R. Epstein, J. Handelsman, G. F. Hatfull, R. Losick, D. K. O’Dowd, B. M. Olivera, S. A. Strobel, G. C. Walker, and M. Warner (2011) Changing the culture of science education at research universities. *Science*. 331(6014): 152-153.
J. Mervis (2013) Transformation Is Possible if a University Really Cares. *Science*. 340 (6130): 292-296.
- 4 J. Fairweather (2008) Linking Evidence and Promising Practices in STEM Undergraduate Education. *National Academy of Sciences (NAS) White Paper*. http://www.nsf.gov/attachments/117803/public/Xc-Linking_Evidence-Fairweather.pdf
A. E. Austin (2011) *Promoting Evidence-Based Change in Undergraduate Science Education*. Paper commissioned by the Board on Science Education of the National Academies National Research Council. Washington, D.C.: The National Academies. http://sites.nationalacademies.org/DBASSE/BOSE/DBASSE_071087#.UdIxQym1F8E

Keynote: Implementing Evidence-Based Undergraduate STEM Teaching Practice

Susan Singer,
National Science Foundation

Developing a diverse, highly skilled, and globally engaged workforce in science, technology, engineering, and mathematics (STEM) and related fields is crucial in advancing the economic and scientific agenda of the nation. Many students who begin their undergraduate years intending to major in a STEM field do not persist, with traditional teaching practices being a common deterrent.¹ Improving the quality of undergraduate STEM education through widespread implementation of evidence-based practices could significantly increase student success and persistence.^{2,3} In the spring of 2012, the release of the President's Council of Advisors for Science and Technology's Engage to Excel report and the National Research Council's synthesis and analysis of evidence-based practices in the *Discipline-based Education Research* report sharpened the focus on implementation of effective practices in undergraduate classrooms across the country.^{2,4} This overview is a synopsis of the efforts that have followed and an exploration of the challenges to widespread uptake of documented approaches to improving student learning, understanding, and persistence in undergraduate STEM fields.

The *Federal Science, Technology, Engineering, and Mathematics (STEM) Education 5-Year Strategic Plan*,⁵ released in 2013, includes an undergraduate education goal to graduate one million additional students with degrees in STEM fields over the next decade by achieving the following strategic objectives:

- 1 Identify and broaden implementation of evidence-based instructional practices and innovations to improve undergraduate learning and retention in STEM and develop a national architecture to improve empirical understanding of how these changes relate to key student outcomes;
- 2 Improve support of STEM education at two-year colleges and create bridges between two- and four-year post-secondary institutions;
- 3 Support and incentivize the development of university-industry partnerships, and partnerships with federally supported entities, to provide relevant and authentic STEM learning and research experiences for undergraduate students, particularly in their first two years; and
- 4 Address the problem of excessively high failure rates in introductory mathematics courses at the undergraduate level to open pathways to more advanced STEM courses.

An interagency working group has been chartered by the National Science and Technology Council's Committee on STEM Education's subcommittee for Federal Coordination in STEM Education with implementing these objectives.

Further, the fiscal year 2014 Cross-Agency Priority Goal for STEM Education specifies implementing the *5-Year Strategic Plan*.⁶

The case for replacing traditional lecture approaches with a broad range of learning experiences documented to enhance learning is compelling. Discipline-centered efforts to leverage this change are growing.^{7,8,9,10,11} Professional societies, including the American Association of Universities (AAU), the Association for Public and Land Grant Universities (APLU), and the American Association of Colleges and Universities (AAC&U) have launched groundbreaking efforts to transform undergraduate classrooms.¹²

We have an effective tool, but many barriers to implementation remain. Faculty time is at a premium and finding solutions that fit into already overly full professional lives is a worthy challenge. Building faculty knowledge of how to use these new approaches in their teaching requires a range of supports. Traditions within the disciplines vary and a one-size-fits-all approach is unlikely to be successful. For many, the content coverage and curriculum remains a stumbling block, though research supports a deep dive into core concepts rather than broad, more superficial content coverage. We continue to need better instruments to assess learning outcomes and understand what is happening as the student moves through the undergraduate years. Building support within departments and institutions is challenging and the support of colleagues is paramount for a faculty member seeking to create a learning environment centered on student engagement in learning. Course scheduling and even space can create barriers. Too often the critical role of non-tenure track and part-time faculty is not fully considered. Student resistance or fear of student resistance can also deter an instructor from easing an evidence-based practice or two into a course.

There are encouraging ways forward, illustrated by the work of the National Science Foundation's (NSF) Widening Implementation and Dissemination of Evidence-based Reforms (WIDER) awardees and the more recent Improving Undergraduate STEM Education (IUSE) program that integrates the WIDER goals. The WIDER solicitation emphasized the importance of beginning with a theory of change for institutional transformation, supported when possible by the literature or designed to add to the knowledge base. Genuine administrative involvement and engagement to improve the chances of an effective outcome was required. Further, establishing baseline data, including the use of evidence-based instructional practices was an essential ingredient. It is difficult to know if approaches to integrating new approaches change outcomes if there is no documented starting point.

Establishing a baseline requires describing and or measuring undergraduate STEM teaching practices. As part of the WIDER initiative, NSF funded an American Association for the Advancement of Science (AAAS) workshop and report on *Describing and Measuring Undergraduate STEM*

Teaching Practices.¹³ As noted in the report, there are several ways to document instructional practices from surveys to interviewing to observation to teaching portfolio analyses. Likewise, these tools can be utilized for different purposes including: non-evaluative improvement of the teaching practice of an individual; tenure and promotion decisions; documenting practices used institutionally or more broadly; and research on teaching and learning. It is important that the same tool not be used for multiple purposes within an institution. Across the nation, the Higher Education Research Institute survey and the National Postsecondary Faculty survey reveal that lecture continues to be the most common teaching modality in STEM classrooms.⁴

A range of experiments is underway to shift instructional practices towards more evidence-based approaches that fully engage students in their own learning. These all push on scaling, within and across institutions. Some build instruments to baseline and assess implementation of new practices. AAU recognized the need for metrics to establish baselines for instructional practices, as well as metrics to assess institutional cultures and impacts of implementation of evidence-based practices as their member institutions sought to move the needle on their teaching cultures.¹⁴ The goal is to develop benchmarks for institutions to measure progress towards reform. At the University of Maine, a classroom dynamics observational protocol is being developed as a tool to engage faculty in 60 STEM courses in professional development to catalyze instructional change.¹⁵ In this case, the instrument is fully integrated into a peer-based professional development program for instructors.

Other tool-based approaches to spreading evidence-based practices focus on supports for busy faculty. Six universities, led by Michigan State, are developing a tool for automated analysis of constructed responses to provide time saving, quality assessments to match the learning outcomes of faculty implementing evidence-based practices.¹⁶ The tool is used in the context of discipline-based professional learning communities.

At the University of Colorado Boulder, Sandra Laursen and her colleagues are engaged in a mixed methods study of a community of mathematicians and mathematics educators dedicated to reforming mathematics education using inquiry-based learning (IBL) mathematics. Laursen's team is clarifying and iteratively examining aspects of a professional learning community that can support and sustain higher-education reform, including both bottom-up and top-down effort to diffuse innovation.¹⁷ While the IBL community is well established, a new community of practice is developing at the University of Illinois, Urbana-Champaign, based on a shift to collective ownership of gateway courses by the faculty, rather than an individual or small group of instructors.¹⁸ Professional development aims to challenge tacit beliefs about teaching, while directly benefitting 17,000 students. Further scaling of professional development is occurring through the Center for the Integration

of Research, Teaching, and Learning (CIRTL), targeting future faculty.¹⁹ Here, MOOCs, including a course on implementing evidence-based practices in the classroom, are underway to make professional development widely available. Further support for faculty in implementing evidence-based practices is anticipated early in 2015 with the release of the National Research Council's practitioner's guide, based on the *Discipline-based Education Research* report.

While the above examples push on faculty knowledge, time savings, and metrics to assess progress, student resistance is also being addressed. The University of Michigan, Ann Arbor; North Carolina A&T; Virginia Tech; and Bucknell University are partnering to understand student resistance to nontraditional teaching methods, informed by social science research on expectancy violation.²⁰ The goal is to develop specific evidence-based strategies that faculty can employ to reduce resistance, thus addressing one of the barriers for faculty implementation.

Given the scale of the challenge we face in widely implementing evidence-based practices, these encouraging approaches have catalytic potential. Each is an opportunity, with carefully designed implementation research, to capture the essential elements, to understand why different elements lead to success, and to then test the most promising approaches in other contexts. The toolkit for changing teaching culture is growing, but fully understanding institutional and system-level transformation remains a worthy aspiration.

- ¹ E. Seymour and N. M. Hewitt (1997) *Talking About Leaving*. Boulder, CO: Westview Press
- ² President's Council of Advisors on Science and Technology (2012) *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*, www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_feb.pdf
- ³ S. F. Freeman, et al (2014) Active learning increases student performance in science, engineering, and mathematics. *PNAS* 111: 8410–8415
- ⁴ National Research Council (2012) *Discipline-based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*. Washington, D.C.: National Academies Press, www.nap.edu/catalog.php?record_id=13362
- ⁵ National Science and Technology Council, Committee on STEM Education (2013) *Federal Science, Technology, Engineering, and Mathematics (STEM) Education 5-Year Strategic Plan*. <http://www.performance.gov/node/3404?view=public#app>
- ⁷ American Association for the Advancement of Science (2011) *Vision and Change in Undergraduate Biology Education*. visionandchange.org/files/2013/11/aaas-VISchange-web1113.pdf
- ⁸ National Research Council (2009) *A New Biology for the 21st Century*. Washington, D.C.: National Academies Press. www.nap.edu/catalog.php?record_id=12764
- ⁹ American Society for Engineering Education (2013) *Transforming Undergraduate Education in Engineering*. www.asee.org/TUEE_PhaseI_WorkshopReport.pdf
- ¹⁰ American Chemical Society (2010) *Chemistry Education: Transforming the Human Elements*. http://archive.aacu.org/pkal/documents/ACS_000.pdf
- ¹¹ National Research Council (2013) *The Mathematical Sciences in 2025*. Washington, D.C.: National Academies Press. www.nap.edu/catalog.php?record_id=15269
- ¹² <http://www.aacu.org/sites/default/files/files/publications/E-PKALSourcebook.pdf>
- ¹³ <http://ccliconference.org/files/2013/11/Measuring-STEM-Teaching-Practices.pdf>
- ¹⁴ http://www.nsf.gov/awardsearch/showAward?AWD_ID=1256221&HistoricalAwards=false
- ¹⁵ http://www.nsf.gov/awardsearch/showAward?AWD_ID=1347577
- ¹⁶ http://www.nsf.gov/awardsearch/showAward?AWD_ID=1347740
- ¹⁷ http://www.nsf.gov/awardsearch/showAward?AWD_ID=1347669
- ¹⁸ http://www.nsf.gov/awardsearch/showAward?AWD_ID=1347722
- ¹⁹ http://www.nsf.gov/awardsearch/showAward?AWD_ID=1347605
- ²⁰ http://www.nsf.gov/awardsearch/showAward?AWD_ID=1347718

Part I: Workshop Sessions

1

Pre- and Post-Testing and Discipline-Based Outcomes

Session Leader: Adam Leibovich, University of Pittsburgh

Panel: Maura Borrego, University of Texas at Austin; Noah Finkelstein, University of Colorado Boulder; Chandralekha Singh, University of Pittsburgh

Summary

It seems almost self-evident that students cannot learn if the teaching we provide is at the wrong level. If the instructor aims too low, the students will be bored and will not learn anything new, while if the instructor aims too high, the students will be frustrated and will not be able to construct an understanding of the material. Theories of learning back up this intuitive idea, such as Piaget’s “optimal mismatch”¹ or Vygotsky’s “zone of proximal development.”² On the other hand, how do we know at the end of the course if the students have actually gained in knowledge? A final exam tests what the students know at the end of the semester, but they may have begun the class with that knowledge. Have the students actually *learned* anything?

Pre- and post-testing solves both of these issues, by giving the instructor knowledge of the students’ “beginning state” and “end state.” Since it is neither time nor resource intensive, pre- and post-testing is perhaps the most straightforward way to introduce reticent instructors to discipline-based educational research validated methods. The workshop covered this in the session entitled *Pre- and Post-Testing and Discipline-Based Outcomes*. Noah Finkelstein, Director of the Center for STEM Learning at the University of Colorado Boulder, gave the motivation for using pre- and post-testing to enhance student learning. Chandralekha Singh, Founding Director of dB-SERC, the Discipline-based Science Education Research Center at the University of Pittsburgh, introduced the theoretical basis of learning and why pre- and post-testing helps close the gap between teaching and learning. Finally, Maura

Borrego, Associate Dean of Engineering at Virginia Tech, explained the nuts-and-bolts on how to use pre- and post-testing and included numerous references to validated instruments.

Noah Finkelstein began the session by emphasizing that it is imperative to professionalize educational practice. Before creating a course, each instructor needs to ask the following questions: What should students learn? What are students actually learning? Which are the appropriate instructional approaches that can improve student learning? Once these are asked, the professor can establish and communicate the learning goals to the class; he or she can deploy appropriate assessment tools and apply research-based teaching techniques. But overall, measuring progress is essential. Assessment allows the instructor to make systematic improvements over time, find and fix problems, and measure impacts on students and instructors. There are many ways we can use these assessments, from assessments aimed at individual lessons to covering the whole course. Frequent formative assessments should be used throughout the course. Finally, Finkelstein emphasized that *how* the assessment is administered matters. For example, it is important for the students to understand what and why they are being assessed, and credit should be given for completing these instruments, but the instructor should not grade the pre- and post-test.

In the next presentation, Chandralekha Singh discussed both a model of student learning and why assessment is crucial to help students evolve from their initial state of knowledge and skills to the final state that the course is designed to convey. Instructors must find out what the students know in order to reach them. Pre-tests need not just be content based. Skills, attitudes, and beliefs can also be covered. The success of instruction depends on the alignment of students' prior knowledge, instructional design, and the evaluation of learning. Singh further emphasized that well-defined course goals and objectives are crucial, and they ought to be explicit. The goal of any course is to increase understanding of the content knowledge, but the student and instructor may have very different notion of understanding. Do we want our students to memorize the definition of acceleration, or to be able to instantiate it in a given situation? Instructors would agree to the latter, but many introductory students believe they understand it if the definition is memorized. Well-defined course goals can then lead to commensurate assessments of learning. Students will acquire usable knowledge and perform well in assessment tasks if they are actively involved in the learning process, have an opportunity to focus on their knowledge structure, and learn from and challenge each other.

Maura Borrego finished the session by outlining best practices for pre- and post-testing. As stressed by the other presenters, first and foremost, the instructor must define outcomes. This ties in to the course goals and objectives for semester long pre- and post-testing, or mini goals for a lesson

plan for shorter-term intervention. When giving any type of assessment, the practitioner should ask whom he or she is trying to convince? Are the assessments being given for the students' sake? Or is it to convince the instructor that the students are learning? Perhaps it is to convince a department chair that the method of teaching is working. Depending on the answer to this question, there are different choices of tools, such as surveys, concept inventories, or standardized tests. Many of these assessments have been validated, and there are places the instructor can go for help in choosing and using these tools, including institutions centers, colleagues, and professional societies. Borrego ended with some cautionary notes on over-interpreting the results of pre- and post-testing.

Action Items

- **Reflect on your teaching.** It is not necessary to become education research scholars, but it is necessary to be scholarly about our teaching. Take time to think about how material should be presented is a start to improved teaching.
- **Use research-based practices.** Try methods already developed and in use. There is no need to develop new methodologies.
- **Add assessments into the course in a thoughtful and consistent manner.** Pre- and post-testing can be incorporated into a course from the start. Pre- and post-testing can help motivate students, measure where students are starting, measure learning gains, and wake faculty up to the learning process.

¹ J. Piaget, *Success and Understanding*, Harvard University Press, 1978.

² L. S. Vygotsky and Michael Cole, *Mind in Society*, Harvard University Press, 1978

2

Student Identification of Learning Outcomes and Improved Student Evaluations

Panel: William R. Dichtel, Cornell University (Panel leader); Susan Elrod, California State University, Fresno; Scott Strobel, Yale University

Summary

Despite the over-reliance on student surveys to evaluate teaching effectiveness, these instruments do provide important insight into student perceptions that should comprise a portfolio of assessment techniques. Surveys should also conform to established best practices to maximize their utility, and all stakeholders must recognize how they should and should not be interpreted based on their known biases and limitations. More broadly, the workshop set out to explore alternative strategies to encourage students to identify learning outcomes, which will increase their ownership, engagement, and autonomy in their education, as well as how to assess such practices.

The workshop addressed these questions in a session entitled “Student Identification of Learning Outcomes and Improved Student Evaluations”. William R. Dichtel, a faculty member in Cornell’s Department of Chemistry and Chemical Biology, described a nationwide survey conducted by the workshop organizers of the current practices to evaluate teaching and learning within STEM disciplines. The survey also measured faculty attitudes towards emerging alternative assessment strategies to gauge the extent of cultural change required for their implementation. Findings of the survey broadly related to student-identified learning outcomes are described below, and a more detailed look at the overall survey results accompanies this piece. Susan Elrod, Dean of the College of Science and Mathematics at Fresno State University, provided her perspective into the composition of student surveys and other assessment strategies needed to inform resource

allocation, tenure and promotion, and program-wide assessment decisions. Finally, Scott Strobel, Henry Ford II Professor of Molecular Biophysics and Biochemistry at Yale University and the Howard Hughes Medical Institute, described a highly successful and engaging laboratory course, in which students travel to Ecuador and discover new endophytes. Although the course is resource intensive, key design criteria were identified that allow this concept to be implemented more broadly, which has been successful in the M2M consortium.¹ Strobel also described an innovative collaboration with David Hanauer from the Indiana University of Pennsylvania that measured student engagement based on a linguistic analysis of their responses to open-ended course survey responses.² This approach appears broadly applicable for measuring student engagement in other laboratory and lecture-based courses throughout STEM programs, and it has since been adapted to evaluate engagement in the Freshman Research Initiative at UT-Austin.³

EETL Group Survey

The Effective Evaluation of Teaching and Learning Survey asked respondents to describe current methods used to evaluate teaching and learning at the course and program level, as well as perceptions of the desirability of several best practices for assessing teaching effectiveness. The survey received 324 responses from chemistry, physics, and astronomy faculty drawn from both primarily undergraduate institutions (PUIs, 67%) and research universities (RUs, 33%), including 38 department chairs. Responses to each survey question were analyzed as a function of institution type (PUI or RU), discipline (chemistry, physics/astronomy), and faculty rank (assistant, associate, full/distinguished professor). As such, we feel that the survey provides key insight into faculty attitudes regarding teaching and evaluation practices. However, we note that the surveyed faculty (other than department chairs) are current and former awardees of grants from Research Corporation for Science Advancement. We suspect that this group is both more likely to use active-learning pedagogies and more open to alternative assessment methods than the average faculty member. As such, learning assessment strategies with limited perceived value in the survey might face major resistance to implementation.

It is sometimes argued that research-intensive universities are less attentive to undergraduate learning outcomes than primarily undergraduate institutions (PUIs). Therefore, we looked for practices used to evaluate teaching and learning at PUIs that might be implemented by Ph.D.-granting institutions. However, we found smaller-than-expected differences in assessment practices between the two types of institutions. Both rely heavily on end-of-semester student evaluations, with PUIs using peer evaluation at a higher rate than research institutions. Subtle differences in practices and attitudes were identified, such as PUI respondents utilizing (and valuing) reflective teaching statements and capstone courses more than their RU

counterparts, but few other assessment tools showed major differences in usage or perceived desirability. It is important to note that our survey does not evaluate the relative merits of undergraduate STEM education in these environments. But it clearly demonstrates that enhancing the assessment of undergraduate learning within Ph.D.-granting institutions will not be as simple as implementing practices already in place at most PUIs.

Not surprisingly, more than 95% of the respondents indicated that their institutions measure teaching quality with paper or online student evaluations. Critics have long argued that the most common forms of these questionnaires are essentially satisfaction surveys that provide little information about actual student learning outcomes.⁴ Although student satisfaction is important, it is only one of several critical metrics that bear witness to student success or teaching efficacy. Of more interest, 50% of institutions use peer observation of in class teaching and a similar number ask faculty to define learning objectives. Reflective teaching statements by faculty are used sometimes (average 1.9 on a 4-point scale), typically more for assistant professors (2.4/4) than full professors (1.5/4). Most other teaching evaluation methods, including teaching portfolios, external review of teaching materials, and rubric-based teaching assessment are currently used rarely.

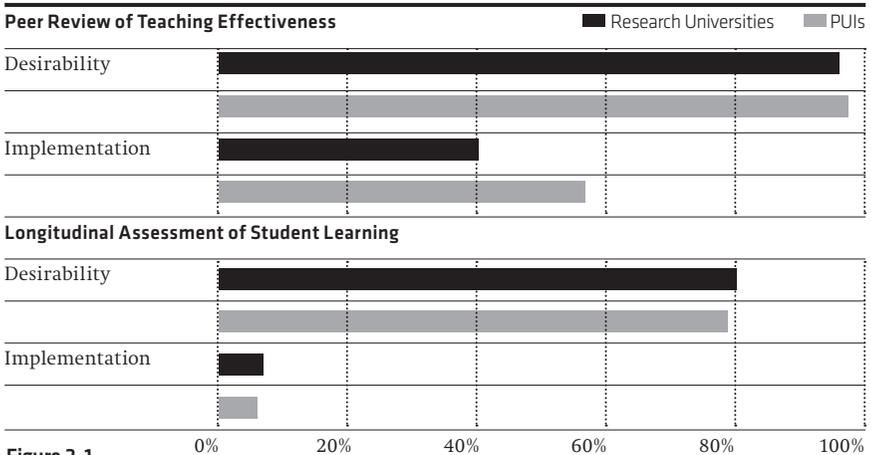


Figure 2-1 Alternative assessment methods are viewed as desirable but are less widely implemented in both research universities and PUIs.

The survey also identified a striking implementation gap in evaluation practices that were perceived to be desirable (Figure 2-1). Greater than 95% felt that peer evaluation of teaching to be highly or moderately desirable, even at RUs, whereas 80% viewed student evaluations favorably. Yet only 50% reported that peer evaluation was used “always” or “regularly” in their departments. Likewise, longitudinal assessment of student learning (i.e., finding out from students how well a pre-requisite class prepared them for a later class) was viewed as ~91% highly or moderately desirable but is used only rarely (~5%

“always” or “regularly”). Assessment methods with large implementation gaps represent excellent opportunities to drive institutional change.

Dean’s Perspective on Improved Learning Assessments

The current over-reliance on student surveys poses significant challenges to university administrators, who must make strategic investments, promotion and tenure decisions, and institutional policy with poor information about the effectiveness of their programs and personnel. Susan Elrod provided a Dean’s perspective into alternative assessment methods that she is implementing at California State University Fresno, and spoke more broadly about how this information can be used effectively and constructively. Elrod posed questions that teaching assessments need to answer, including:

- “How does an instructor define their teaching job?”
- “Are students learning and progressing? Are they engaged in the learning process?”
- “Are courses modern and up-to-date? Do they leverage faculty research expertise?”
- “How does what this faculty member is doing contribute to the program? Should the faculty member be retained or tenured?”
- “Are we delivering the program we advertise? Is it the best we can offer?”

Elrod described a hierarchical approach to evaluate programs in the context of the above questions. After defining a program-level learning outcome (e.g., what students should learn in a major), she evaluates ways in which students will demonstrate mastery of these concepts and how the program will assess this mastery. Finally, the contributions of individual courses and faculty members, as well as how specific pedagogies are implemented, may be judged in the context of how they contribute to the program. Student survey data remains valuable to these assessments, and Elrod advocated that they be modified to incorporate aspects of learning-assessment instruments, such as Student Identification of Learning Gains (SALG).

In evaluating courses and faculty members, a high priority was placed on reflective teaching practices throughout the workshop. Elrod advocated for faculty to prepare course portfolios, which describe formal course goals and outcomes, strategies to meet these goals, and evidence of student learning. Elrod also emphasized that course portfolios are a flexible platform that might be required or voluntary. Portfolios have gained traction elsewhere—they are now required at the University of Arizona and are sometimes used in judging campus-wide teaching awards. Nevertheless, they may face resistance to broader implementation, as nearly one-third of our survey respondents (and half of the department chairs) currently perceive them as “not valuable” or “too much effort for the possible gains”.

Reflective practices are also important at the departmental level to evaluate degree programs. Degree programs should offer synergy among their courses, as well as connections and reinforcement of important

concepts across multiple courses, such that a STEM degree involves deeper mastery than the sum of the learning outcomes of its individual courses. Elrod proposed a skill inventory model to evaluate student-learning outcomes across a major. Skills identified by the faculty as particularly important for their field should be developed across multiple courses and communicated clearly to the students.

Discovery-Based Undergraduate Research Experiences and Assessing Student Engagement

Discovery-based research courses challenge students with open-ended problems and encourage educational engagement. These courses give students control over the direction of a project with an uncertain outcome and sufficient flexibility to be taken in many different directions. Recognizing their potential, the President's Council of Advisors on Science and Technology (PCAST) report strongly recommended that they replace traditional laboratory courses in STEM curricula.⁵

Strobel described his course, "Rainforest Expedition and Laboratory" at Yale, in which students isolate and characterize plant-associated fungi known as endophytes from Yasuni National Forest in Ecuador. The majority of these endophytes are unknown, providing students ample opportunity to isolate new species and characterize their function and biochemistry. The course, which spans the spring semester and following summer, is split into three phases: project development, sample collection in Ecuador during Yale's spring break, and laboratory work for identification and chemical analysis (Figure 2-2). Undergraduates enrolled in the course have discovered many previously unknown endocytetes, averaging one new fungal genus per student year, and have sometimes isolated bioactive compounds produced by these fungi.

Students in the Rainforest Expedition course developed a strong sense of project ownership and engagement compared to those in standard laboratory

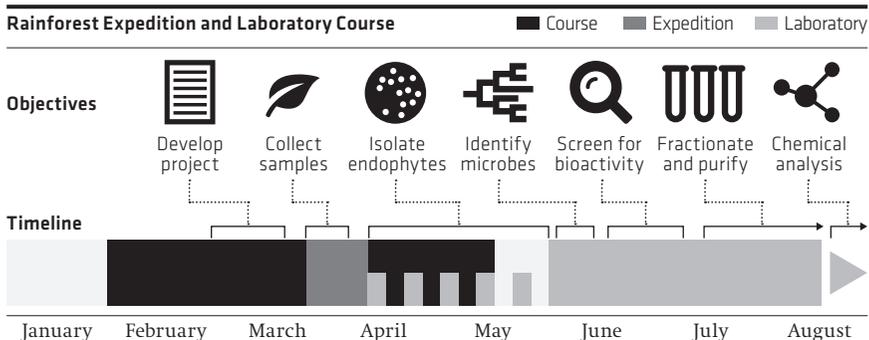


Figure 2-2

Approximate timeline for the Rainforest Expedition and Laboratory Course.

courses. Strobel and Prof. David Hanauer, a linguist from Indiana University of Pennsylvania, quantified this effect by analyzing the content and linguistic structure of student responses to exit interview questions. For example, content analysis indicated that students in the Yale course expressed greater sense of personal scientific achievement than those enrolled in either independent research or standard laboratory courses. The rainforest expedition students also were more likely to use personal pronouns and emotional words in their exit interview responses.⁶ These findings support the drive to implement research-based courses more broadly. They also suggest that linguistic analysis of open-ended student survey questions might be more broadly applied to evaluate student engagement in non-laboratory courses or degree programs.

It is important to note that successful student engagement in research-based courses has also been achieved in other innovative programs, such as: UT-Austin's Freshman Research Initiative;⁷ Georgia Tech's Vertically-Integrated Projects (VIP) Program;⁸ BioLEd at the University of Virginia;⁹ and Consortium to Promote Reflection in Engineering Education at the University of Washington.¹⁰ Together with the Yale course, these programs serve as models for institutions with differing resources and enrollments. In general, discovery-based courses must give students control over a project that has an uncertain outcome. Strobel also emphasized that the project should be large and sufficiently flexible that students play an active role in determining its direction.

Action Items

- **Implement alternative teaching and learning assessments.** Departments and administrators must be cognizant of the limitations of student surveys, particularly when boiling teaching effectiveness down to a single numerical measure judged by students. This practice is inaccurate and actually causes faculty to be risk-averse regarding new teaching pedagogies. Alternative assessments should be chosen to best fit the needs and resources of each institution or program. Several alternative assessments already have significant faculty buy-in but are not practiced widely—including peer observation of teaching.
- **Best practices for student surveys:** Student surveys should be retained, but should include questions regarding learning-based assessments, not just student perceptions. For example, faculty should identify specific learning objectives at the beginning of the course, and the surveys should ask students whether they felt that they learned them.
- **Adopt alternative teaching and learning evaluations:** Our survey of faculty perceptions identified an implementation gap for some evaluation methods perceived to be useful, most notably peer evaluation of teaching. Such methods are promising in that they face the smallest hurdles to achieve faculty and

administrative buy-in. Each institution, or even individual degree programs, should adopt alternative evaluation techniques that they determine to be both implementable and effective; there is no one-size-fits-all approach.

- **Encourage and assess student engagement:** Laboratories provide an outstanding and often untapped opportunity for students to take ownership of their own learning. Such opportunities should be tailored to the resources of different academic programs and institutions, but should move beyond canned, outcome-defined laboratory experiments. Assessment of these exercises should measure student engagement. The linguistic analysis pioneered by Hanauer is an intriguing option that might prove broadly applicable.

¹ <http://cst.yale.edu/M2M>

² D. I. Hanauer, J. Frederick, B. Fotinakes, S. A. Strobel (2012) "Linguistic analysis of project ownership for undergraduate research experiences." *CBE-LSE* 11, 378-378.

³ D. I. Hanauer, E. L. Dolan (2014) "The project ownership survey: measuring differences in scientific inquiry experiences," *CBE-LSE* 3,149-158.

⁴ S.R. Porter (2011) Do College Student Surveys Have Any Validity? *The Review of Higher Education*, 35, 45-76.

A. Greenwald (1997) Validity concerns and usefulness of student ratings of instruction. *American Psychologist* 52, 1182-1186.

"An Evaluation of Course Evaluations" (2014) 10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1

⁵ President's Council of Advisors on Science and Technology (2012) "Engage to Excel: Producing One Million Additional College Graduates With Degrees In Science, Technology, Engineering, And Mathematics." http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_feb.pdf

⁶ D. I. Hanauer, J. Frederick, B. Fotinakes, S. A. Strobel (2012) "Linguistic analysis of project ownership for undergraduate research experiences." *CBE-LSE* 11, 378-378

⁷ <http://cns.utexas.edu/fri/about-fri>

⁸ <http://vip.gatech.edu/>

⁹ <http://biochemlab.org>

¹⁰ <http://depts.washington.edu/celtweb/cpree/>

3

Peer Observation and Evidence of Learning

Session Leader: Andrew Feig, Wayne State University

Panel: Gail Burd, University of Arizona;

Pratibha Varma-Nelson, Indiana University—Purdue University of Indianapolis;
and Robin Wright, University of Minnesota

Summary

The third session focused on peer observation of teaching. In advance of the workshop, a survey was conducted of faculty and department chairs that focused on the current mechanisms used to evaluate faculty teaching and the desirability of other methods that might not be in current use. That survey is described in more detail in Chapter 2. Based on the survey responses, speakers were asked several questions to help frame the discussions:

- Should we observe teaching and if so, by whom?
- Are there dangers or threats to the use of peer observation?
- What training should observers undergo to prepare them for the task?
- Observation is nearly universally accepted as a formative process, but should it also be used for summative assessment (i.e. promotion and tenure)?
- Are there biases associated with peer observation?
- How should observation data be used and who should have access to it?

Speakers were chosen to represent several different perspectives on the peer observation discussion. One speaker serves as director of a center for teaching and learning, a second is a senior vice provost responsible for academic affairs and the third, an associate dean who oversees teaching and undergraduate issues within a college of biological sciences.

Pratibha Varma-Nelson, Director of the Center for Teaching and Learning (CTL) at IUPUI and professor of chemistry, discussed the role of Centers for Teaching and Learning in institutional programs for peer observation of teaching. She identified several critical elements for the success of these

mechanisms and summarized the literature on best practices for peer observation. Two important lessons came out of this discussion. The first was that you need to clearly articulate the goal of the observation before setting in place specific observation protocols. This ensures that the practices that are used for the observation lead to the desired outcomes, especially if the goal is teaching reform, programmatic alignment or preparation of materials for promotion and tenure. In general, CTLs are willing to help train peer observers, but try to keep distanced from summative assessments. Since the CTLs primary mission is faculty professional development through formative consultation, they need to maintain this distance from P&T evaluation.

With the shift toward more online courses, a critical question has emerged regarding how such courses should be evaluated? In many cases, the course design is done by one person (or team of people) while the actual implementation of the class any given term is someone else entirely. Thus, some of the elements of teaching are outside the control of the instructor.

Understandably, as a director of a CTL, it was the speaker's view that CTLs are effectively performing their primary function of supporting the teaching mission of colleges and universities. Not all institutions have such centers however and there are questions of scalability to a wide array of institutions. Especially when funding is tight, they may not receive the financial support they need to fulfill their missions.

The second question that arose pertains to who uses these facilities and services and how CTLs can be used more effectively. If CTLs are viewed as places to send faculty to be "fixed" then the center becomes undermined. Instead, CTLs must become a nexus for discussions on effective teaching and a voice on campus for the celebration of exemplary teaching. Finding mechanisms through which a greater proportion of the faculty interact with CTLs to continue their own education on how to be effective instructors is central to their mission. Thus, for reflective instructors who are looking to improve their own efficacy, the CTL becomes an important ally on campus.

Gail Burd, Senior Vice Provost at University of Arizona, Distinguished Professor of Molecular and Cellular Biology, and coordinator of their UA AAU STEM Education Project, spoke about the institutional changes in the promotion and tenure procedures associated with their AAU program. The primary gist of the revised protocols provides greater emphasis on the teaching component of the *research:teaching:service* formula. Some of these changes include the use of teaching portfolios and teaching observations using Classroom Observation Protocol for Undergraduate STEM (COPUS) as documentation of teaching efforts and approaches.¹ While UA and a number of other institutions have adopted the COPUS protocols for STEM teaching observation due to its ease of implementation for P&T, the University of Arizona employs a UA designed observation protocol that can be tailored to specific teaching approaches and methods as specified by the

instructor. As with IUPUI, the center for teaching and learning performs only formative consultations with faculty. Summative assessments of faculty for P&T are gathered by peer faculty observers from the department of the faculty member.

The survey of faculty performed in advance of the workshop (see Chapter 2) showed that faculty respondents were overwhelmingly unsupportive of the use of teaching portfolios for documentation of teaching excellence. While this can obviously be mandated by the administration, it will be interesting to assess whether University of Arizona faculty are more supportive of portfolios in a few years after having experienced their use.

This talk and the discussion that ensued touched on a potential third-rail issue—that of the university reward structure and the rewriting of P&T guidelines to provide for a more prominent role for instruction. Some of the concerns however have been how to provide a more fair and consistent approach to the evaluation of teaching excellence. Research is more easily quantified by STEM departments. There are clear units of scholarship and quantifiable income from external grants that mark a “return on investment” relative to startup funds invested. Many STEM departments are less comfortable with the quantifiable parameters associated with faculty teaching and student learning. Thus the ability to demonstrate the reliability of certain measures like COPUS observations and their link to student performance is still important to document. In the meantime, however, the revised University of Arizona guidelines indicate a clear shift in priorities that stipulate the value added that exceptional teachers bring to the university.

Robin Wright, Associate Dean of Biological Sciences at University of Minnesota provided the third perspective on the issue of peer evaluation of teaching. Her take on the issue was significantly different than the other viewpoints. In collaboration with the Vision and Change movement, she has been a champion of breaking down the perception that teaching is a private activity between a faculty member and their students. Her view is that if these privacy walls can be overcome, then the need for teaching observation by definition goes away because teaching becomes a communal activity—visible to any and all interested stakeholders who wish to observe and participate in the conversation about teaching and learning. The question is how to make such a vision come to be given current faculty attitudes.

At Minnesota, they are pursuing this cultural shift by creating comfortable spaces to share teaching. The first approach involves team teaching assignments. The novelty here is the manner in which these shared teaching assignments are orchestrated. While most schools use serial team teaching, the Minnesota approach involves both faculty being present at all class sessions so that the faculty are active collaborators in the teaching process. This is particularly useful for aligning overall programs and ensuring curricular continuity when teaching assignments transition to new

instructors. In addition, it is used in a very intentional way to mentor faculty. This can be used for new faculty who are just developing their teaching style, or with faculty who want to make a concerted change in their approach, such as those who want to transition from lecture-based teaching to active learning strategies. In this way, new implementers of these methods have a way to learn from a faculty member already using the methods or to get real-time assistance in the management of such discussions.

The second approach in use at the University of Minnesota is a faculty learning community that meets regularly over lunch to talk about teaching and related topics. These gatherings provide a forum where faculty with common interest in student achievement gather to share notes. This sends a strong message that teaching is part of the common mission of the faculty and sets the stage for sharing that in most departments is limited to research issues.

What comes out of this discussion is a potential scalable implementation that might move departments toward more open teaching practices. Implementation requires the assistance and cooperation of the dean, however, who typically would control the course load assignments. By using release time for faculty mentoring activities, a dean can make a strategic investment in either a junior faculty member's development into a reflective teacher or a senior faculty members exploration of new teaching methods. This will only be effective however, if it is clear that the time cannot be used for serial teaching by two faculty, but instead is true team teaching where both faculty are present at every class session and work together to achieve student-centered outcomes. The faculty member with the release time can then also be tasked with the administrative role of overseeing the faculty learning community with the department.

The discussion after the three formal talks focused on several issues, but the most notable was probably that of **incentivizing and empowering faculty to work together for the improvement of teaching practices**.

Action Items

- **Expanded role and impact for Centers of Teaching and Learning.** Overall, the discussion illustrated that Centers for Teaching and Learning are serving an important function in the professional development of faculty through the process of formative consultations. However, centers for teaching and learning can be much more significant partners in the transformation of instructional practices. This requires that faculty champions partner with these centers as liaisons and agents of transformation. Taking advantage of faculty with the disciplinary credentials to bridge the gap between regular faculty and teaching specialists is critical to gaining broad-based buy in for the uptake of evidence-based teaching.

- **Co-teaching and teaching mentoring.** Departments regularly assign new faculty research mentors who assist assistant professors make the transition from post-doc to successful faculty member. We support the concept that faculty be assigned teaching mentors as well. The University of Minnesota approach of using team teaching is one way to implement this. Teaching mentors would receive workload credit for their service in this capacity, jointly teaching a course to provide direct interaction between a master educator and a junior faculty member. Similarly, dual teaching in this way can be used during the transition between an incoming and outgoing instructor in large service courses to provide continuity and ensure the sustainability of reforms as teaching assignments shift over time.
- **Promotion and tenure.** So long as tenure review is dominated by research performance above all else, teaching will remain a distant secondary concern to research productivity for junior faculty. Formal review criteria must mirror the departmental and institutional priorities and represent the manner by which such values communicated to the faculty and tenure committees.
- **Periodic department and program review.** While a lot of the discussion focused on individual teaching activities, the discussion also dealt with issues of the overall interrelation of teaching within a department or major. Individual teaching is important, but so is alignment of the curriculum and the mapping of critical concepts to individual courses within the major. The best time for such plans to be reviewed, revised and evaluated is as part of the departmental review. Typically, program reviews focus on research productivity of the faculty and graduate program health. We propose that administrators and department chairs acknowledge the importance of undergraduate teaching as part of these reviews by ensuring that at least one member of the review committee be a specialist in undergraduate teaching within the discipline.

¹ M. K. Smith, F. H. M. Jones, S. L. Gilbert, and C. E. Wieman, The Classroom Observation Protocol for Undergraduate STEM (COPUS): A New Instrument to Characterize University STEM Classroom Practices, *CBE-LSE* 2013;12:618-627

4

Analytics and Longitudinal Assessment

Session Leader: Steve Bradforth, University of Southern California

Panel: Steve Benton, IDEA Center; Marco Molinaro, University of California, Davis; and Lynne Molter, Swarthmore College

Summary

Although evidence-based learning has become a mantra of higher education, universities themselves do a relatively poor job of collecting evidence of quality in classroom teaching and fail to learn much from the data they do collect. There is a particular problem in getting useful measures of performance and student learning back to faculty. Everyone participating at this workshop agreed that good quality data, particularly longitudinal assessment of faculty teaching in early courses and its impact in student performance later in the degree program, has the potential to drive change at many levels. This session discussed three working models for improving educational outcomes by looking at where students come from, collecting data based on targeted questions of both students and faculty and digging deep into the data to understand how students' performance in STEM courses reflects classroom teaching practices. These models range from more careful evaluation of how student perceptions of a class compare with faculty-defined learning objectives, to a recent pilot research project tracking student—particularly minority student—migration through majors and finally to a broad university initiative which makes heavy use of data analytics. Analytics is a technique widely deployed in business to inspect large and complex datasets and by transformation or modeling, discovering useful information that suggests relationships and support decision-making.

The IDEA Center Student Rating of Instruction, described by Steve Benton, an IDEA Senior Research Officer and Emeritus Professor of Education

Psychology at Kansas State, rests on the idea that specific teaching methods influence certain types of student learning; their project has years of data from hundreds of institutions to draw from. The CUSTEMS pilot study described by its co-creator Lynne Molter, Professor of Engineering at Swarthmore College, uses data collection from student arrival on campus through graduation to identify longitudinal patterns of migration into and out of STEM majors. The iAMSTEM Hub project, a million-dollar investment by the Provost's office at UC Davis headed up by Marco Molinaro, both collects and aggregates fine-grained information on student performance and intervenes to provide solutions at a department and individual faculty level. Molinaro is director of the two-year-old hub and Assistant Vice Provost for Undergraduate Education at UC Davis.

In his presentation, Benton quickly pointed out that a student is not qualified to evaluate either the faculty's teaching or in fact their own learning.¹ So one might ask why do university administrators then use student surveys as the primary means to *assess* and *evaluate*? Nevertheless, because student evaluations of teaching (SET) are based on the observations of a large base of raters and across multiple occasions, they are statistically the most reliable measure of an instructor's effectiveness. With the rather generic questions surveyed in SETs, it is still controversial precisely what measures of effectiveness are actually being gauged and it is influenced by many other factors than learning.² But if best practices regarding question design are followed, IDEA argues that student responses do illuminate their perceptions of the learning expectations and outcomes of a course. The IDEA survey instrument triangulates whether the student perceptions match with the instructor's expected learning objectives. For example, compared to non-STEM classes, STEM instructors in introductory classes emphasize different learning objectives, particularly cognitive and application objectives. Even when a STEM instructor includes other learning objectives (*e.g.*, critical analysis of ideas, team exercises, acquiring an interest in learning more) in their responses students recognize making less progress in these areas. The IDEA surveys also found that students taking STEM classes to satisfy general education or distribution requirements gain the ability to apply course materials and develop better core competencies when "the instructor frequently inspires students to set and achieve goals that really challenge them."

Funded by the Sloan Foundation, the Consortium for Undergraduate STEM Success (CUSTEMS) is a pilot project aiming to address issues related to undergraduate degree completion in STEM fields, with a particular focus on under-represented minority students. One clear goal is to identify high-risk students early. Molter described the set of metrics (from parent's ZIP code, to Math SAT and college gateway STEM grade) and the cluster analysis their collaboration had established to identify the reasons for student retention,

migration and attrition. In practice, the students are surveyed in their first semester both for anticipated major as well as to measure demographic information. A fourth semester survey then looks back at their choices, links to surveys as well as registration data. The CUSTEMS data shows that net migration into STEM subjects is lower than into humanities, and there is a lower retention overall of women than men in STEM fields. The range of thirty institutions represented in the consortium is quite varied and each had committed to adopting common strategies for successful educational outcomes for STEM students.

Molter emphasized that this longitudinal analysis should be used to identify and address systemic problems in a degree program, rather than to identify or predict outcomes for individuals based on their demographic information. It was noted that there is a danger of using analytics predictively at the student advising level. What tends to then happen is a perpetuation of your observations, for example a higher drop out rate of a given under-represented group.

Marco Molinaro described a multifaceted project at UC Davis that collects detailed data for each course (student performance and satisfaction along with the teaching practices of individual instructors) at every stage along a student's trajectory and then analyzes overall outcomes. This approach is applied for students on a single campus right across the disciplines of agriculture, medicine, science, technology, engineering and math (the iAMSTEM Hub). The project, which has received a million-dollar investment from UC Davis has a clear goal to improve undergraduate STEM student success. One of their strategies is to build rich analytic tools and include these in a comprehensive framework to measure and inform on improvement of student outcomes and associated teaching practices. Rather than using analytics approaches to the learning process itself⁹, the Hub collects classroom observation information, aggregates student admission, performance and SET data from multiple university repositories and tries to break down roadblocks to getting information to faculty they can use to refine their teaching. Molinaro also was quick to point out that this analytic information should not be used to predict individual performance. It is generally true that in large introductory STEM classes, instructors genuinely don't know much about the students they are teaching. The Hub can however flag simple trends— for example, when an instructor consistently grades 0.75 GPA points higher leading to students waiting on spaces in this instructor's section to open up.

Within the data gathered, longitudinal trends are established, but testing somewhat different factors to the CUSTEMS project. The iAMSTEM Hub team asks whether there is a correlation in the point at which a student gives up on a major, or a pre-professional track, with the taking of a course from a specific faculty member at a key juncture. In principle, data can be used to answer whether a given sequence properly prepares a student for an upper-

division class. A novel aspect of this project is the close relationship between the Hub and contributing department heads as well as individual faculty. Solutions were offered to faculty rather than simply pointing out problems. A component that was widely appreciated at the workshop was the powerful visualization tools that have been developed as part of the Hub. For example, ribbon visualizations powerfully illustrated longitudinal aspects in STEM retention: where students who start in a given major end up. It was widely commented that the visualization tools developed in this project could be readily adopted at other institutions. The model of having an experienced STEM educator, rather than an institutional research office, interpret the analytics and communicate the results with faculty, was also seen as a highly desirable aspect that should be replicated in implementation elsewhere.”

Action Items

- **Keying in on course learning objectives.** Student Evaluations should ask students to self-report their progress against instructional objectives defined by the professor at the beginning of a course. This a likely indicator of their achievement of course learning outcomes.
- **Importance of a broad range of assessment inputs.** Each of the presenters stressed that the collection of data, even the more incisive information analyzed by these approaches, should only form one part, perhaps no more than 50%, of the overall evaluations of teaching and other measures should be included such as classroom observation of practice, pre- and post- testing of learning gains, etc.
- **Broaden adoption of analytics and longitudinal approaches.** The iAMSTEM Hub is recognized to be an enormously powerful model, although the cost and logistics of setting up an equivalent program at other institutions is likely to be cost-prohibitive for most. However, many of the tools that have been developed at UC Davis are already being freely shared via a consortium and much of the experience learned can be applied at other institutions.
- **Institutional configuration of longitudinal data analysis.** To harness data from a multitude of campus databases, the person heading up an analytics function must have an appointment, or full support, of the provost. However, unlike staff in the traditional institutional research office, it is key to appoint an individual who has the connections and rapport with the STEM departments and direct experience in STEM teaching issues, including an understanding of patterns of success for students from different socio-economic background.

- ¹ S. L. Benton and W. E. Cashin (2012) Student Ratings of Teaching: A Summary of Research and Literature. *IDEA Technical Report #50*, Manhattan: Kansas State University, Center for Faculty Evaluation and Development.
- ² K. A. Feldman (1997) Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry and J. C. Smart (eds.), *Effective teaching in higher education: Research and practice* (pp. 368-395). New York: Agathon Press
- H. W. Marsh and L. A. Roche (1997) Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52(11), 1187-1197.
- R. Sproule (2000). Student evaluation of teaching: A methodological critique of evaluation practices. *Education Policy Analysis Archives*. <http://epaa.asu.edu/ojs/article/view/441> (accessed May 25, 2015).
- ³ M. C. Wright, T. McKay, C. Hershock, K. Miller, and J. Tritz (2014) Better Than Expected: Using Learning Analytics to Promote Student Success in Gateway Science. *Change: The Magazine of Higher Learning* 46(1):28-34. DOI: 10.1080/00091383.2014.867209

5

Administration and Implementation: Incentivizing, Uses, and Abuses of Evaluation and Assessment

Session Leader: Emily Miller, Association of American Universities

Panel: Karen Bjorkman, University of Toledo; Kathryn Miller, Washington University in St. Louis; and Mary Ann Rankin, University of Maryland, College Park

Summary

Universities and colleges are complex environments in which many factors facilitate, impede, or influence change. Relying on her well-documented systems approach to change, A. E. Austin (2011) stipulates that sustainable change in undergraduate STEM education requires an understanding of the overall system in which undergraduate education is situated. Single-lever strategies are unlikely to result in transformation in undergraduate STEM learning. Successful transformation efforts require multiple facilitators or “levers” pushing for change to counterbalance the forces that sustain ineffective instructional practices (Austin, 2011).

Institutional leaders such as department chairs, deans, and provosts can facilitate or impede transformation of undergraduate STEM education. During the session, *Administration and Implementation: Incentivizing, Uses, and Abuses of Evaluation and Assessment*, we heard from institutional leaders at each of these three administrative levels.

Kathryn Miller serves as the Chair, Department of Biology at Washington University in St. Louis, Karen Bjorkman serves as Dean of the College of Natural Sciences and Mathematics at the University of Toledo, and Mary Ann Rankin, Provost at University of Maryland, College Park. Miller drew upon her institutional and national leadership roles to improve teaching methodology and to reform STEM curriculum. The Partnership for Undergraduate Life Sciences Education (PULSE) selected Miller as one of 40 *Vision and Change Leadership Fellows* tasked with improving undergraduate life-sciences education

nationwide. As a former Chair of the Department of Physics and Astronomy and current Dean, Bjorkman commented on her approach to promoting a culture that values teaching. Rankin reflected on her experience as Dean of the College of Liberal Arts and Sciences at UT Austin and her leadership of a number of institution-wide projects to improve undergraduate STEM teaching and learning such as, Influence of Discovery Learning Program, UTeach, and the Freshman Research Initiative.

Cross-cutting all three speakers were the common themes that institutional leaders must provide a coherent message, support faculty professional development, and allocate resources to improve undergraduate STEM teaching. In addition, they impact institutional reward processes that constitute a major driving force for effecting faculty members decisions regarding teaching and time allocation to diverse job duties.

Department chairs, deans and provosts provide symbolic support that signals that good teaching is valued and rewarded. They shape the extent of faculty members' attention to teaching, the choice of particular teaching and learning practices, and readiness for or resistance to innovation in teaching. "Teaching needs to be part of conversations in a meaningful way," said Bjorkman.

Faculty members often are unfamiliar with the research on learning and teaching and how to implement evidence-based teaching practices. Department chairs can initiate departmental conversations around curriculum and assessment. In addition, as noted by Miller, department chairs can provide growth-oriented, non-remedial, time-effective, accessible, and individually relevant professional development for faculty members.

Whether faculty work assignments provide for time spent on teaching improvement is also an institutional factor related to change. Department chairs are in the position to provide faculty the time to learn and implement new approaches to teaching. As Rankin reflected, "once you get faculty engaged in new teaching paradigms that work—everyone wants to be a good teacher but not all know how. The level of excitement, no matter what the reward structure, is palpable."

Institutional leaders need to allocate specific resources to achieve defined teaching goals and be patient during periods of change. Faculty members are penalized in the early stages of adopting new teaching practices is detrimental to achieving cultural change. For example, at the departmental level, administrators can promote teaching arrangements where faculty pairs work in partnership to improve teaching practices by providing both faculty members workload credit. At the institutional level, one can incorporate into fundraising campaigns support for endowed chairs of teaching that provide recognition for teaching innovation or empower faculty members to invest in professional development activities to redesign their courses and corresponding teaching practices.

Additionally, with strong institutional commitment, it is possible to implement campus-wide analytics that utilizes existing student learning and outcome data. Harnessing longitudinal data on student progress has the power to inform discussions and decision making about teaching and learning at the individual faculty member, department and institutional levels. Data, when used effectively, provides the continuous feedback that allows rapid realignment of teaching methodologies and implementations to inform institutional change and engage faculty in the conversation about student learning.

In order to achieve real and lasting change, all three panelists emphasized that colleges and departments will need to begin to utilize alternative metrics of teaching performance to supplement the degree to which student evaluations are currently used to assess teaching performance. Rankin stated that “rigorous and enforced promotion requirements that included teaching excellence,” is necessary. Yet even if these alternative methods are developed and implemented broadly, there is still another step that must be taken. This involves additional training and encouragement for members of promotion and tenure review committees to: 1) effectively evaluate and assess teaching quality even if such additional measure are provided to them, and 2) give teaching effectiveness appropriate weight when compared to research in such P&T decisions.

There is no one-size-fits-all solution, and progress will likely occur more slowly than many of us would prefer. Nevertheless, given improved knowledge about which STEM teaching practices are the most effective, as well as growing external pressures to improve the quality of undergraduate STEM education at research universities, change is imperative. Moving forward will require a commitment at all levels—upper level administration, colleges, departments, and faculty—to share and adapt the practices to overcome the inherent obstacles to systemic and sustainable change in undergraduate STEM education.

As a first step to removing these long-standing impediments, institutions and their colleges and departments must expect and enable their faculty members to be scholarly about their teaching. They must also support, assess, recognize, and reward those who are. Thoughtful analytics and new tools that evaluate student learning and teaching quality at both the faculty and the departmental level present a powerful new mechanism for accomplishing this goal. If applied, these new tools will lay the path toward more balanced recognition of STEM teaching and research at our universities.

Action Items

- **Establish a culture of conversations about undergraduate teaching and learning within STEM departments.** Senior university administrators play critical roles in promoting a culture that values teaching and a support structure that fosters continuing improvement and innovation.
- **Provide faculty with adequate time, resources, and professional development to improve teaching.** Teaching assignments should accommodate the challenges and time commitment necessary to improve pedagogies and integrate new learning assessment techniques. Departments and colleges should encourage faculty members and teaching assistants to utilize their campus centers for teaching and learning, which help foster a culture of continuous teaching improvement.
- **Provide centralized and accessible data and analytics.** Universities accumulate volumes of longitudinal data that have been underused in assessments of student learning and degree progress. Administrators must implement a robust, scalable, and centralized campus-wide analytics approach that uses existing data and reduces the need to create multiple assessment tools. Once data are collected and analyzed, the findings can be shared with departments and faculty and used constructively and formatively.
- **Use teaching improvement as a fundraising lever.** Alumni, as well as private and public sector employers, have direct interests in enhancing the university teaching and learning experience. Senior administrators should seek support for well-articulated initiatives to improve STEM education by incorporating them into their fundraising campaigns.
- **Set a positive tone by recognizing and rewarding good teaching.** University administrators influence how faculty members prioritize their teaching and research efforts. Colleges and departments must do more than pay lip service to the role of teaching in annual review, contract renewal, and promotion and tenure processes.

Part II: Submitted Papers

6

Pre- and Post-Testing in STEM Courses

Maura Borrego,
The University of Texas at Austin

A pre/post test design involves administering the same or similar assessment measures before and after an intervention, to assess the impact of that intervention. It is important to define desired outcomes in clear statements of what you want students to get out of the course or intervention. In addition to pre/post assessment designs, you may also want to consider alternative designs, depending on who they are trying to convince. Quantitative pre/post data is useful if trying to convince managers, funders, or peer reviewers; post-test only or qualitative designs may suffice if you just need to convince yourself. Examples of tools that might be used as pre- and post-tests include: surveys, concept inventories, standardized tests (to measure critical thinking, cognitive ability, attitudes), and rubrics to analyze student products.

Concept inventories are a particularly popular tool in STEM education. They are a series of multiple choice, conceptual questions (formulas, calculations, or problem solving skills not required) with the possible answers including distracters representing common student misconceptions. Published concept inventories are extensively validated; developing one is a considerable undertaking. There are several published instruments for characterizing students' intellectual development: Measure of Intellectual Development (MID), Measure of Epistemological Reflection (MER), Learning Environment Preferences (LEP), and the Critical Thinking Assessment Test (CAT).

Rubrics are typically more context-specific and do not necessarily require as much validation work. Rubrics are written scoring tools that indicate various levels of quality for achieving the goals of the assignment. Rubrics are

often shared with students when an assignment is made, to communicate expectations and grading schemes. Rubrics work particularly well for communication, portfolio, and project assignments. Scores derived from rubrics may be more convincing if an outsider (expert) scores anonymous student work.

There are several resources for finding validated assessment instruments:

- STEM specific: <http://assess.tidee.org/>
- STEM Concept Inventories: <https://cihub.org/>
- ETS Testlink
- Mental Measurements Yearbook
- Tests in Print
- Literature searches with keywords: survey, inventory, questionnaire, evaluation, assessment

You should work with your university librarian to access these and other databases, and to refine your searches. Once you identify a tool for potential use in your pre/post test, consider the nature of the tool (whether it really fits what you are interested in), prior validation of the tool (statistics and prior work with a population similar to your students), and experience others have had with the tool (costs, availability of comparison data). If you ever want to publish your results, for example as a conference paper, seek IRB approval in advance. This will probably require you to use codes or another identifier (not students' names) to match the pre- and post-tests to each other for statistical analysis.

Students should be given ample time to complete the tests so that they measure learning rather than efficiency. Some assessment experts recommend that they be separate from course grades, or students can receive points for completing the assessment but not a score based on their answers. Although online administration has clear benefits, the response rate is much higher for in-person paper tests. When interpreting your pre/post test results, consider alternative explanations. For example, if you observed a change, students may have learned the concept outside of class, other changes in the course could have caused the improvement, the pre-test or post-test data could be distorted by an external event, or the instrument may be unreliable. If no change is observed, it could be due to sloppy administration of the instrument, non-random drop-out between pre- and post-tests, faulty analysis, lack of motivation of respondents, or the intervention not being effective or not long enough. An external evaluator or other collaborator can assist with pre/post testing and interpreting the results. You can find an evaluator or collaborator at your institution's Center for Teaching and Learning; Center for Evaluation; colleges and departments including Education, Ed Psych, Engineering, Science, Public Admin; colleagues; professional societies; or the American Evaluation Association web site. Your evaluator will bring to the collaboration:

knowledge about evaluation design and methodology, experience in evaluating your type of project, a quality evaluation plan, a contract or an agreement with deliverables and timeline, knowledge of evaluation standards, and frequent, respectful communication. Your evaluator will expect you to provide a clear definition of your project goals, objectives, activities, and expected outcomes; all available information about your project; frequent communication; questions; access to data/participants; help with problem-solving; and resources and payment.

Resources

Felder, R. M., S. D. Sheppard, and K. A. Smith. 2005. A new journal for a field in transition. *Journal of Engineering Education* 94:7–10

NSF's Evaluation Handbook. Available at: <http://informal.science.org/documents/TheUserFriendlyGuide.pdf>

Olds, B. M., B. M. Moskal, and R. L. Miller. 2005. Assessment in engineering education: Evolution, approaches and collaborations. *Journal of Engineering Education* 94:13–25

7

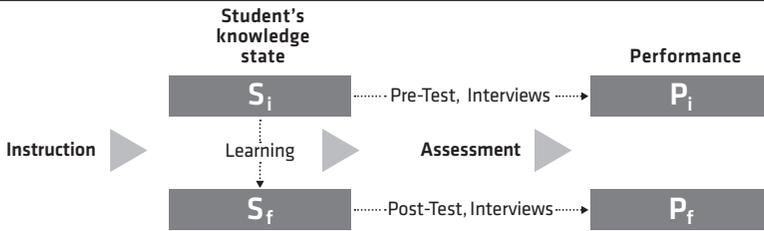
Closing the Gap Between Teaching and Assessment

Chandralekha Singh,
University of Pittsburgh

Evidence-based teaching is based upon a model of learning in which assessment plays a central role.¹ New knowledge necessarily builds on prior knowledge. According to the model shown schematically in Figure 7-1, students enroll in a course with some initial knowledge relevant for the course. The instruction should be designed carefully to build on the initial knowledge and take students from that initial knowledge state to a final knowledge state based upon the course goals. It is difficult to know what students know at the beginning or at the end of a course. Assessment is the process in which both students and instructor get feedback on what students have learned and what is their level of understanding vis a vis the course goals. To assess learning, students can be given pre-tests and post-tests at the beginning and at the end of the course. The performance on these tests can reflect the extent to which the course goals have been achieved if the tests are designed carefully consistent with the course goals to reflect students' knowledge state accurately.

Research suggests that students must construct their own understanding though the instructor plays a critical role in helping students accomplish this task. An instructor should model the criteria of good performance, while leaving sufficient time to provide guidance and feedback to students as they practice useful skills. The amount of instructional support given to students should be decreased gradually as they develop self-reliance. A particularly effective approach is to let students work in small groups and take advantage of each other's strengths.

Evidence-Based Teaching


Figure 7-1

Model of learning in which assessment plays a central role.

Discipline-based education research emphasizes the importance of having well-defined learning goals and assessing student learning using tools that are commensurate with those goals. “Students should understand acceleration” is not a well-defined goal because it does not make it clear to students what they should be able to do. Students may misinterpret this instructor’s goal to imply that they should know the definition of acceleration as the rate of change of velocity with time, while the instructor may expect students to be able to calculate the acceleration when the initial and final velocities of an object and the elapsed time are provided. Examples of measurable goals should be shared with students and should include cognitive achievement at all levels of Bloom’s taxonomy including students demonstrating how to apply concepts in diverse situations, analyzing problems by breaking them down into sub-problems, synthesizing a solution by combining different principles, and comparing and contrasting various concepts.

Assessment drives learning and students focus on learning what they are tested on. Therefore, using assessment tools that only probe mastery of algorithms and plug and chug approaches will eliminate incentives to acquire deep understanding. Furthermore, instructional design should be targeted at a level where students struggle appropriately and stay engaged in the learning process and the material should not be so unfamiliar and advanced that students become frustrated and disengage. Research-based curricula and pedagogies are designed to take into account the initial knowledge of a typical student and gradually build on it.

The course assessments should be viewed as formative rather than summative. If used frequently throughout the course, formative assessments can greatly improve the knowledge students have about the course material, since they have many opportunities to reflect on their learning consistent with course goals. In addition, frequent assessment can help the instructor get real time feedback on the effectiveness of instruction which can be used to refine instruction and address difficulties.

Formative assessment, *e.g.*, think-pair-share activities, clicker questions,

tutorials with pre- and post-tests, collaborative problem solving, process oriented guided inquiry learning, minute papers, asking students to summarize what they learned in each class or asking them to make concept maps of related concepts etc. should be used throughout the course in order to help students build a good knowledge structure and develop useful skills. Using these low-stakes built-in formative assessment tools throughout the course can bridge the gap between teaching and assessment.

To summarize, all of the formative assessment approaches fulfill at least in part the following criteria:

- They provide students with an understanding of the goals of the course because the activities that they engage in communicate instructor's expectations (I expect that you are able to solve this type of problems, complete these types of tasks etc.)
- They provide students with feedback on where their understanding is at a given time with relation to the course goals as communicated by the instructor
- They provide the instructor with feedback on where the class is with relation to the course goals
- They make students active in the learning process and students obtain timely feedback which helps them improve their understanding early when there is time to catch up

These examples suggest an important aspect of formative assessment: significant use of formative assessment tools which are carefully embedded and integrated in the instruction entails a close scrutiny of all aspects of an instructional design. Before implementing evidence-based teaching and learning one should compile a list of what initial knowledge students have, what measurable goals are intended to be achieved via the course and think carefully about how to design instruction aligned with the initial knowledge of students and course goals and how to scientifically assess the extent to which each goal is achieved. In essence, evidence-based teaching entails that instructors carefully contemplate the answers to the following questions:

1. What is the initial knowledge of the students that is relevant for instruction (content-based initial knowledge including alternative-conceptions, problem solving and reasoning skills, mathematical skills, epistemological beliefs, attitude, motivation, self-efficacy etc.)?
2. What should students know and be able to do?
3. What does proficiency in various components of this course look like?
4. What evidence would I accept as demonstrating proficiency? What evidence would be acceptable to most of my colleagues?
5. How can I design instruction that builds on students' initial knowledge and takes them systematically to a final knowledge state which is commensurate with course goals?
6. Are the initial knowledge of students, course goals, instructional design and assessment methods aligned with each other?

A typical goal of a science course is to provide students with a firm conceptual understanding of the underlying knowledge. Therefore, discipline-based

education researchers in many disciplines have developed assessment instruments designed to assess students' conceptual understanding. The data from these instruments can be used to assess the initial knowledge of the students if administered as a pre-test before instruction and inform what students learned and what aspects of the material were challenging if given after instruction as a post-test. These data can help improve instructional design, *e.g.*, by pinpointing where more attention should be focused to improve student learning. If the post-test is administered right after instruction in a particular course, it can serve as a formative assessment tool and if it is given at the end of the term, it can still be helpful for improving learning for future students.

Another course goal may be to improve students' attitudes about the nature of science and learning science and provide them with an understanding of what science is and what it takes to be successful in science courses. To this end as well, discipline-based educational researchers have developed assessment instruments that are either discipline specific or about science in general. Also, in many science courses, students are expected to develop good problem solving strategies and to this end, instruments have been developed to assess students' attitudes and approaches to problem solving.

The standardized assessment instruments that have been developed for science courses provide a starting point for thinking about assessing effectiveness of teaching and learning and investigating the extent to which various instructional goals have been met. The data obtained from these instruments can be compared with national norms found in publications in education research journals. See <http://www.dbserc.pitt.edu/> for examples of such instruments in natural sciences, *e.g.*, physics, chemistry, biology, mathematics etc.

¹ National Research Council, *Knowing what students know: The science and design of educational assessment*, Committee on the Foundations of Assessment, eds. J. Pellegrino and R. Glaser, National Academy Press, Washington, D.C., 20001

8

Improving Teaching Quality through Peer Review of Teaching

Gail D. Burd, Molecular and Cellular Biology, Office of the Provost, University of Arizona; Ingrid Novodvorsky, Office of Instruction and Assessment, University of Arizona; Debra Tomanek, Molecular and Cellular Biology, Office of Instruction and Assessment, University of Arizona; and Pratibha Varma-Nelson, Department of Chemistry and Chemical Biology, Center for Teaching and Learning, Indiana University-Purdue University Indianapolis

Introduction

There are no shortcuts if proper evaluation of teaching is to be accomplished. Several mechanisms are used to evaluate teaching quality. These include:

- teacher-course evaluations by students in the course,
- focus group interviews with faculty and students,
- self-survey of faculty teaching practices,
- student learning gains,
- classroom observations of teaching practices,
- peer observations of teaching for formative and summative evaluations,
- team teaching with instructional feedback, and
- portfolios assembled by faculty to document teaching quality through description of teaching practices and innovations, students learning outcomes, peer observations of teaching, and faculty teaching awards.

No one mechanism is perfect by itself, but collectively, several different sources of information can provide a clear picture of teaching quality.

At the University of Arizona in January 2014, a survey was given to 73 registered participants in a workshop entitled “Peer Evaluation of Teaching” with the question “What do you currently do to receive feedback on the quality of your teaching?” The responses were: 1) 85% use teacher-course evaluations by students, 2) 14% report that they only use peer reviews, and presumably do not use student evaluation or do not think the student evaluations provide feedback on teaching quality, 3) 36% use some peer review or team teacher feedback, and 4) 8% use graduate teaching assistant feedback. In response to the question “Do you believe that having peer review

of your teaching is useful?”, most answered that they thought peer review would be valuable. Some respondents, however, replied that it would depend on who did the review, how the review was done (*e.g.*, using a rubric), and for what purpose. Assuming that the University of Arizona survey results reflect practices at other AAU institutions it is clear that colleges and universities need to think carefully about how best to collect information on the quality of teaching.

Because so many faculty use teacher-course evaluations, we need to explore the pros and cons of student evaluations of the course and the teacher(s) as an assessment of teaching quality. In a 1997 review of research on student evaluations of teaching effectiveness, Marsh and Roche state “students’ evaluations of teaching (SETs) are (a) multidimensional; (b) reliable and stable; (c) primarily a function of the instructor who teaches a course rather than the course that is taught; (d) relatively valid against a variety of indicators of effective teaching; (e) relatively unaffected by a variety of variables hypothesized as potential biases (*e.g.*, grading leniency, class size, workload, prior subject interest); and (f) useful in improving teaching effectiveness when SETs are coupled with appropriate consultation.” The authors also indicate that the teacher-course evaluations are a sufficient assessment tool for effective teaching when combined with other sources of information. Faculty, departments, and institutions use student evaluations because they are relatively easy to do, give the students a voice in the evaluation process, and provide quantitative data important for comparisons with other faculty and for a single faculty member over time. Students, however, lack the experience to evaluate fully the content knowledge of the instructor and the value of the concepts selected for the class. Furthermore, students may be unaware of teaching strategies that their instructor could use to more fully support their learning. Therefore, we suggest that peer review is one way to address some of the shortcomings associated with using SETs as the only form of evidence when evaluating teaching quality.

For our discussion in this session entitled “Peer Observation and Evidence of Learning,” several questions were posed by the session leader, Andrew Feig, to the speakers and participants.

- Should we use peer review of teaching?
- Who should be selected to provide peer observations?
- What training is needed to help observers provide effective feedback?
- What observer biases might exist?
- Are there dangers or threats caused by the observations?
- How can institutions use teaching observation data?
- Should institutions do summative assessment? And if yes, at what frequency?

In this chapter, we address several of these questions.

Quality Teaching and Student Learning

The most important reason to evaluate teaching is to improve the quality of teaching and to enhance student learning. Thus, the best approaches will be those that give confidential, formative feedback to the faculty member about their teaching along with measurements of student learning outcomes (discussed in other chapters in this volume).

What characterizes good teaching? Chickering and Gamson (1987) indicate that a good teacher encourages contact between the student and the instructor, provides opportunities for active learning, helps develop cooperation among students, gives prompt feedback to students, emphasizes time on task, communicates high expectations, and respects diverse talents and learning styles. Wieman (2012) states “effective teaching is that which maximizes the learner’s engagement in cognitive processes that are necessary to develop expertise.” A good analogy might be that a good teacher can be considered a personal trainer or coach, someone who will challenge the student to achieve his or her potential through motivation and active-learning with deliberate practice (Wieman, 2012).

Learning can be defined as information, ideas, and skills that a person can use after a significant period of disuse and can apply to a new problem. Assessment of student learning outcomes needs to address what students can do with what they learned. It is not enough to have memorized a series of facts; students need to develop deep understanding of the concepts and be able to apply that knowledge, or transfer the knowledge, to novel situations and problems. (Bransford *et al.*, 1999)

Evidence-Based Teaching Practices that Lead to Enhanced Learning

As documented in the NRC report “Discipline-based education research: understanding and improving learning in undergraduate science and engineering” by Singer, Neilson, and Schweigruher (2012), teaching practices that are based on research and that include active and student centered learning are much more effective than the traditional method of lecturing. Several classroom observation instruments have been developed over the past few years that align with the practices reported in the 2012 NRC report. A number of these instruments are reviewed in the AAAS report “Describing & Measuring Undergraduate STEM Teaching Practices” (2012) (see Table 8-1).

Table 8-1: Classroom Observation Instruments**Classroom Observation Protocol for Undergraduate STEM (COPUS)**

<http://www.cwsei.ubc.ca/resources/COPUS.htm>

Reformed Teaching Observation Protocol (RTOP)

https://mathed.asu.edu/instruments/rtop/RTOP_Reference_Manual.pdf

UTeach Observation Protocol (UTOP)

http://cwalkington.com/UTOP_Paper_2011.pdf

Oregon Collaborative for Excellence in the Preparation of Teachers Classroom Observation Protocol (OTOP)

<http://opas.ous.edu/Work2009-2011/InClass/OTOP%20Instrument%20Numeric%202007.pdf>

Teaching Behaviors Inventory (TBI)

<http://www.calvin.edu/admin/provost/documents/behaviors.pdf>

Flanders Interaction Analysis (FIA)

E. J. Amidon and N. A. Flanders (1967) *The role of the teacher in the classroom. A manual for understanding and improving teachers' classroom behavior.* Minneapolis: Association for Productive Teaching.

Teaching Dimensions Observation Protocol (TDOP)

<http://tdop.wceruw.org/>

VaNTH Observation System (VOS)

<http://onlinelibrary.wiley.com/doi/10.1002/j.2168-9830.2003.tb00777.x/pdf>

A. H. Harris and M. F. Cox (2003) Developing an Observational System to Capture Instructional Differences in Engineering Classrooms. *J. Engineering Education*. 92:329-336.

Classroom Observation Rubric

C. Turpen and N. D. Finkelstein (2009) Not all interactive engagement is the same: Variation in physics professors' implementation of peer instruction. *Physics Review Special Topics: Physics Education*, 5:020101.

Not yet available for undergraduate STEM education observations at the time the AAAS report was published, but now in use by several research universities, is the COPUS. The COPUS protocol of Smith and colleagues (2013) does not include mechanisms for positive or negative evaluative input; it is just a straight observation tool predicated on the idea that active-learning instructional approaches lead to increased learning. Yet, when combined with evidence of student learning outcomes, the COPUS can serve as a valuable tool for culture change toward evidence-based instructional practices. For example, the University of Arizona recently combined student learning outcomes with COPUS observations. In this preliminary study of a physics class, the investigators showed that an instructor who uses evidence-based, student-centered teaching practices compared to a traditional lecturer can provide compelling data to departmental faculty about the value of active-learning instructional practices. In this case, the traditional lecturer was convinced of his need to change his teaching practices. The University of Arizona is now beginning to modify the COPUS to incorporate evaluation of the quality of the classroom practices that are observed.

The COPUS observation protocol requires a moderate amount of training, which may make it difficult for most departmental faculty. However, individuals in centers or offices of teaching and learning can provide trained professionals or, as in use now at the University of California at Davis, train teaching assistants to perform the observations.

Survey tools and observation protocols that indicate the extent to which evidence-based teaching methods are used can serve as a proxy for increased student learning. Thus, the teaching practices inventory by faculty (Wieman and Gilbert, 2014) <http://www.cwsei.ubc.ca/resources/TeachingPracticesInventory.htm> and <http://www.lifescied.org/content/13/3/552.full.pdf+html> and the Classroom Observation Protocol for Undergraduate STEM (COPUS) (Smith, *et al.*, 2013) <http://www.cwsei.ubc.ca/resources/COPUS.htm> and <http://www.lifescied.org/content/12/4/618.full> are built on the premise that teaching and learning are improved by the use of active-learning instructional practices. Wieman has suggested that if an institution were to do one thing to assess and improve teaching, it should start with understanding the teaching practices of the faculty (pers. comm.).

The advantages to the teaching practices inventory (2014) are that no training is needed, and the survey only takes about ten minutes to complete. Furthermore, the answers can be used to measure the extent to which the instructor uses active-learning instructional practices and a score can be awarded based on the answers to the survey questions. The limitation is that results of the survey may not be a factual account of the teaching practices of the instructor since the results are self-reported.

Protocols for Peer Observations

At institutions of the authors of this chapter, central facilities on teaching and learning provide support for classroom observations. Indiana University-Purdue University Indianapolis (IUPUI) has professionals who will make classroom observations and provide feedback to instructors (<http://ctl.iupui.edu/Services/Classroom-Observations>). Instructors who have requested an observation meet with a consultant prior to the classroom visit or online course review to clarify the goals of the observation and review course materials. Following the observation, the faculty member and consultant create teaching improvement goals for which strategies are selected, designed, and implemented to meet. The Center for Teaching and Learning at IUPUI also offers workshops for departments upon request so they can design their own peer review procedures. This is important to get faculty buy in. Training for individual faculty or groups of faculty to conduct effective peer reviews is also available.

The University of Minnesota Center for Teaching and Learning <http://www1.umn.edu/ohr/teachlearn/resources/peer/index.html> developed peer review protocols that can assist individuals, departments, or programs

through the following steps of the process of peer review by:

- helping departments establish and implement a peer-review process,
- helping departments improve their current peer-review process,
- preparing individuals to participate in the peer-review process by helping them document their teaching and compile the appropriate materials,
- preparing individuals to carry out a peer review of their colleagues, and
- providing examples of peer review systems.

The University of Arizona also recently developed an interactive website with a protocol that can be used for formative or summative peer observations of teaching <http://oia.arizona.edu/project/peer-review-teaching-protocol>.

Table 8-2: Topics for UA Classroom Observation Tool

1	Lesson Organization
2	Content Knowledge
3	Presentation
4	Instructor-Student Interactions
5	Collaborative Learning Activities
6	Lesson Implementation
7	Instructional Materials
8	Student Responses

The UA template for peer observation of teaching includes several observation items under each topic and can be filled out online during the observation. The website provides suggestions for the review process, offers a menu of observation items, and leads to an interactive rubric that can be used at the time of the observation. During a pre-observation meeting, the instructor and observer select the review topics and observation items that will be used. Little or no training is needed to use the protocol. It has been designed to allow senior faculty within a department or program to provide formative or summative review of teaching for teaching improvement and for annual evaluations and the promotion and tenure review of teaching. Since it is not possible to review teaching on all the topics available for peer observation, it is suggested that departments agree which of the observations topics and items will be used in summative evaluation of teaching.

Another excellent source for peer review of teaching is provided by Chism (2007). The rationale and approach to the peer review process is provided along with suggested questions that can be used to design protocols for peer review of specific types of teaching. This volume also presents ideas for making discussions about teaching more open among colleagues within in departments.

Formative Peer Observations of Teaching

Formative peer observation is used for improvement of teaching. This should be done in a way that is not threatening to faculty. Offices that provide faculty teaching support often have professionals who can give feedback on pedagogy, teaching practices, instructional approaches and assessment, and often are able to observe faculty while they are teaching to provide formative evaluation. Since these offices are most often focused on assisting faculty with instruction and assessment and are used for professional development, they should not be used for peer observations that lead to evaluations for promotion and tenure or annual evaluations. Asking these professionals to both support and evaluate faculty sets up an uneasy tension and can lead to suspicion and mistrust on the part of the faculty.

In addition, since professionals in offices of teaching and learning are most often not in the discipline of the faculty member they are observing, they will likely not be able to comment on course content. Colleagues in the same or a very similar field, however, can provide feedback on the selection of course content, course organization, course objectives, course materials, student evaluation measures, methodology used to teach specific content areas, teaching practices that enable students to learn in the field, and teaching practices used by effective teachers in the field.

Feedback on the last two areas may depend on the teaching philosophy and teaching practices used by the peer evaluator. Therefore it is best to get input from several sources.

This leads to the question of who should provide the peer evaluation of teaching and how this person is selected. Formative evaluation is peer observation designed to be used to improve teaching, and information from the observation should go only to the faculty member. The faculty member who would like to improve his/her teaching practices and course materials, organization, and assessment methods should select someone they perceive to be an excellent teacher. This may be a faculty member who has won teaching awards or someone the faculty member has observed teaching and wants to emulate.

Table 8-3: Recommendations for Formative Peer Review of Teaching Using the UA Peer Observation Protocol

1	Pre-observation meeting to discuss the target class and goals for the observation
	<ul style="list-style-type: none"> • Instructor provides overview of the course • Instructor outlines what will take place during the observed class period(s) • Instructor indicates the learning goals of the lesson(s) • Instructor discusses the type of feedback he/she hopes to receive • Instructor and observer select the topics and items from the observation tool
2	Observer visits the target class(es), completes the Classroom Observation Tool, and prepares a written summary of the observation
3	Post-observation meeting to discuss the observed class(es)
	<ul style="list-style-type: none"> • Observer asks the instructor what he/she thinks worked well in the lesson • Observer asks the instructor what he/she thinks could have been improved • Observer comments on selected items from the Classroom Observation Tool. The selected items may include: <ul style="list-style-type: none"> a organizational skills and instructional approaches observed during the class period, b clarity of the instructions and responses to questions, c apparent attitude of the students and their time on task during the class period, and d summary or closure of the lesson at the end of the class. • Feedback should be: <ul style="list-style-type: none"> a focused on improvements, b non-judgemental, c offer constructive suggestions as options, d action oriented, and e given in ways that have the instructor develop ownership of the ideas.
4	This cycle could lead to another class observation by the same observer to provide further feedback on any changes that the instructor made after the previous observations.

* <http://oia.arizona.edu/project/peer-review-teaching-protocol>

Another, more informal, but effective approach to formative peer observation that is used at the University of Minnesota is team teaching. In addition, the University of North Carolina at Chapel Hill has developed a formal co-teaching program that includes a mentor and an apprentice faculty member, and the mentor is not always the senior of the two instructors. Teams that include a distinguished instructor with someone who wants to improve their teaching can be very effective. Teaching teams that share the teaching activities together on a daily basis can provide regular peer review of teaching. The distinguished instructor can serve as model for the novice and the novice can practice new teaching approaches under the guidance of the distinguished instructor.

What are the costs and benefits of using teaching teams with a novice and an experienced teacher? The benefits include:

- the ability of the experienced instructor to model how the lesson content can be presented to increase learning,
- the experienced member of the teaching team can provide regular feedback to the novice instructor on teaching and thus reduce stress that may come from being observed on a rare occasion, and
- instructors in teaching teams often find it rewarding to share and try new ideas for teaching.

To be effective and provide the benefits listed above, both instructors should be in the classroom together for most lessons. This, however, can lead to significant challenges related to:

- time and workload for the instructors, and
- cost to the department by assigning two faculty to one class for most class sessions.

Summative Peer Observations of Teaching

Unlike formative peer observation of teaching, summative peer observations are designed to provide an evaluation of the faculty member's teaching that will go to the department head, promotion committee, or in other ways be used for promotion and tenure. The faculty, department head, and dean need to agree in advance of any evaluations what aspects of teaching will be included in peer observations for annual evaluations and for promotion and tenure reviews. They should also approve an observation rubric that will be used for all teaching evaluations in the department. It is critical that these reviews be fair, consistent and reliable across different observers, and compatible with accepted standards of good teaching. Furthermore, there should be more than one observation, observations should cover the full class period, a consistent and approved rubric should be used during the observation, the observation should be preceded by a pre-observation meeting and followed by a post-observation meeting, and the faculty member should receive a copy of the observation report. Ideally, two or more faculty reviewers will participate in the observations, but this can create workload challenges for a small department.

Institutions have a need for summative evaluation of teaching quality. The selection of the peer faculty member(s) who will provide summative evaluations for promotion and tenure requires careful consideration by the department chair. Often the department chair will want to select a senior faculty member for this task. However, young faculty may be more innovative and more likely to use evidence-based teaching practices, while senior faculty may use traditional lecturing in their teaching. Challenges may arise if the selected evaluator is unfamiliar with current teaching practices that use active-learning instructional practices; the class can seem chaotic or unfocused to a traditional lecturer making the observation.

At the University of Arizona, peer review of teaching is now a required

component of the tenure package. At other colleges and universities, this summative peer evaluation of teaching may take place annually for all faculty or only for the untenured faculty, at a pre-tenure review (most often three years into an assistant professor position), the year before the tenure decision, the year before any faculty promotion, or as part of a required teaching improvement plan.

Formative Review and Summative Evaluation of Teaching

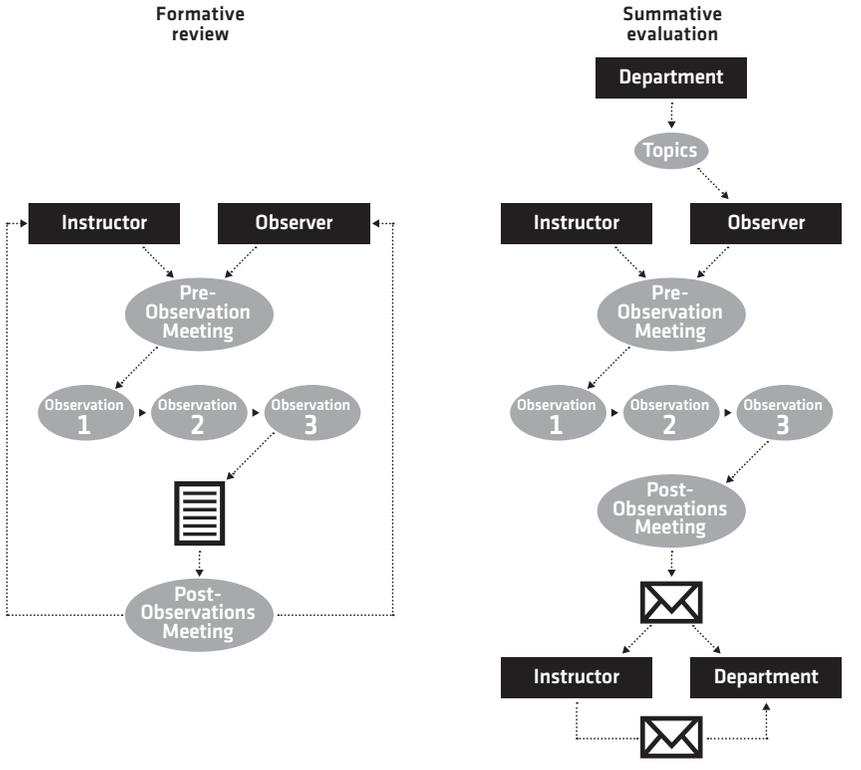


Figure 8-1

Illustrates the differences between formation review and summative evaluation of teaching. The participants shown in the illustration are: instructor, observer, department and instructor. The activities include a pre-observation meeting, three observations, a post-observations meeting, department. Documents include topics for review selected by faculty in department, a memo and a letter. The departmental review process is described above.

Summary

In summary, to get a balanced assessment of teaching quality, institutions should use a variety of evaluation approaches. Student evaluations can be valuable, but provide an incomplete picture of a faculty member's teaching effectiveness. Knowledge about teaching practices used in the classroom can be obtained through the teaching practices inventory (Gilbert and Wieman, 2014) or peer observations such as the COPUS (Smith *et al.*, 2013). Use of teaching practices that actively engage students in constructing their own knowledge have been shown to improve learning outcomes (Wieman, 2012). Therefore, knowing what teaching practices are used provides some evidence about teaching quality. Formative peer review of teaching through observations and follow-up conversation are the most valuable mechanism for improving teaching quality. Selecting the most appropriate peer observer and using pre- and post-observation meetings and validated rubrics can improve the feedback provided to the instructor. Team teaching with an instructor who uses active-learning instructional practices can facilitate informal formative feedback on teaching quality. Summative peer reviews of teaching have a role to play in colleges and universities for annual and promotion and tenure evaluations, but the observer should be selected with care and the topics for review and observation need to be consistent and approved by departmental faculty. Peer observation of teaching and teaching materials along with assessment of student learning outcomes can also be used for program assessment and improvement.

Nearly all of the current science education literature points to the value of active-learning instructional practices and student centered learning that require the student to be more engaged. This is the teaching approach we want STEM faculty to use. Faculty who rely primarily on lecture for transmitting knowledge are encouraging students to memorize material for the test, but evidence shows that learning through memorization is not retained. On the other hand, faculty who use teaching practices that are evidence-based encourage students to gain conceptual understanding of the course material and to be able to transfer that knowledge to novel problems in the future. Whatever procedures we use to provide peer observation of teaching, we need to keep our focus on improving students' learning outcomes.

References

- AAAS Report. (2012). *Describing and Measuring Undergraduate STEM Teaching Practices*. AAAS, Washington, D.C.
- Amidon, E. J. and N. A. Flanders. (1967). *The role of the teacher in the classroom. A manual for understanding and improving teachers' classroom behavior*. Minneapolis: Association for Productive Teaching.
- Bransford, J. D., A. L. Brown, and R. R. Cocking, eds. (1999). *How People Learn: Brain, Mind, Experience, and School*. National Academy of Sciences. Washington, D.C.
- Chickering, A. W. and Z. Gamson. (1987). Seven principles of good practice in undergraduate education. *Amer. Assoc. Higher Ed. Bulletin*, pp. 3-7.
- Chism, N. V. N. (2007). *Peer Review of Teaching: A Sourcebook (2nd Edition)*, Bolton, MA: Anker
- Harris, A. H. and M. F. Cox. (2003) Developing an Observational System to Capture Instructional Differences in Engineering Classrooms. *J. Engineering Education*. 92:329-336.
- Marsh, H. W. and L. A. Roche. (1997). Making students' evaluations of teaching effectiveness effective. *Amer. Psychologist* 52:1187-1197.
- Singer, S. R., N. R. Nielsen, H. A. Schweingruber. (2012). *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*. National Academies Press, Washington, D.C.
- Smith, M. K., F. H. M. Jones, S. L. Gilbert and C. E. Wieman. (2013). The classroom observation protocol for undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. *CBE-LSE* 12:618-627.
- Thomas, S., Q. T. Chie, M. Abraham. (2014). A qualitative review of literature on peer review of teaching in higher education: An application of SWOT Framework. *Rev. Ed. Research* 84:112-159.
- Turpen, C. and N. D. Finkelstein. (2009). Not all interactive engagement is the same: Variation in physics professors' implementation of peer instruction. *Physics Review Special Topics: Physics Education*, 5:020101.
- Wieman, C. and S. Gilbert. (2014). The teaching practices inventory: a new tool for the evaluation and improvement of college and university teaching in mathematics and science. *CBE-LSE* 13: 552-569.
- Wieman, C. (2012). Applying new research to improve science education. *Issues in Science Technology*. pp. 1-7. <http://issues.org/29-1/car1/>

9

Student Ratings of Instruction in Lower-Level Postsecondary STEM Classes

Steve Benton,
Senior Research Officer, IDEA Education, and Emeritus Professor,
Educational Psychology, Kansas State University

When used appropriately, student ratings of instruction can be a valid and reliable indirect measure of student perceptions of how much they have learned in a course and of how effectively the instructor taught (Benton & Cashin, in press; Benton & Cashin, 2012; Marsh, 2007; McKeachie, 1979). Although faculty and administrators often refer to them as “course evaluations” or “student evaluations,” *student ratings of instruction* is the preferred term. “Evaluation” refers to a determination of worth and requires judgment informed by multiple sources of evidence. “Ratings” are but one form of data that require interpretation. Using the term “ratings” helps to distinguish between the source of information (student perceptions) and the process of assessing value (evaluation) (Benton & Cashin, in press). In the end, student ratings should count no more than 30-50 percent of the overall evaluation of teaching.

Student ratings are, nonetheless, the most reliable single measure of teaching effectiveness because they are based on the observations of multiple raters across multiple occasions. Their validity is supported by evidence of positive correlations with other relevant criteria, including instructor self-ratings, ratings by administrators and colleagues, ratings by alumni, ratings by trained observers, and student achievement (Benton & Cashin, in press). However, no single measure of teaching effectiveness is sufficient. Ratings should be combined with other indicators (*e.g.*, peer observations, administrator ratings, alumni ratings, course materials, student products). Moreover, students are not qualified to rate some important aspects of teaching, such as subject-matter knowledge, course design, curriculum

development, commitment to teaching, goals and content of the course, quality of tests and assignments, and so forth. (For a review of these and other issues, see Abrami, d'Apollonia, & Rosenfeld, 2007; Abrami, Rosenfeld, & Dedic, 2007; Arreola, 2006; Braskamp & Ory, 1994; Benton & Cashin, 2012; Cashin, 1989; Cashin, 2003; Centra, 1993; Davis, 2009; Forsyth, 2003; Hativa, 2013a, 2013b; Marsh, 2007; Svinicki & McKeachie, 2011).

Researchers should select a student ratings system backed by reliability and validity evidence and comprised of standardized items and scores. One such system is IDEA Student Ratings of Instruction (SRI) (<http://ideaedu.org/>). Developed by faculty at Kansas State University who had won teaching awards, and supported by a Kellogg Grant awarded in 1975, IDEA is one of the oldest and most widely used systems in higher education.

Since its inception, IDEA rests on its student-learning model, which holds that specific teaching methods influence certain types of student progress (i.e., learning) under certain circumstances. Faculty rate each of 12 learning objectives (see Table 9-1) as *Essential*, *Important*, or *Of Minor or No Importance*.

Table 9-1: IDEA Learning Objectives

1	Gaining factual knowledge (trends, etc.)
2	Learning fundamental principles, generalizations, or theories
3	Learning to apply course material (to improve thinking, problem solving, and decisions)
4	Developing specific skills, competencies, and points of view needed by professionals
5	Acquiring skills in working as a team member
6	Developing creative capacities (writing, art, etc.)
7	Gaining a broader understanding and appreciation of intellectual/cultural activity (music, science, literature, etc.)
8	Developing skill in expressing oneself orally or in writing
9	Learning how to find and use resources
10	Developing a clearer understanding of, and commitment to, personal values
11	Learning to analyze and critically evaluate ideas
12	Acquiring an interest in learning more

Students, in turn, rate their progress on the same 12 objectives, using the scale *No apparent progress* (1), *Slight progress* (2), *Moderate progress* (3), *Substantial progress* (4), or *Exceptional progress* (5). IDEA statistically adjusts student progress ratings on relevant objectives (those the instructor rated as essential or important) for class size and student ratings of their work habits and desire to take the course (i.e., course circumstances beyond the instructor's control). Students also indicate how frequently their instructor used each of 20 teaching methods by responding *Hardly Ever* (1), *Occasionally* (2), *Sometimes* (3), *Frequently* (4), or *Almost Always* (5). The teaching methods (see Table 9-2) are conceptually tied to Chickering's and Gamson's (1987) principles of good practice and are grouped into five underlying teaching styles based on factor analysis (Hoyt & Lee, 2002a).

Table 8-2: Teaching Method Styles on the IDEA Student Ratings Diagnostic Form

I	<p>Stimulating Student Interest</p> <ul style="list-style-type: none"> 4 Demonstrated the importance and significance of the subject matter 8 Stimulated students to intellectual effort beyond that required by most courses 13 Introduced stimulating ideas about the subject 15 Inspired students to set and achieve goals which really challenged them
II	<p>Fostering Student Collaboration</p> <ul style="list-style-type: none"> 5 Formed “teams” or “discussion groups” to facilitate learning 16 Asked students to share ideas and experiences with others whose backgrounds and viewpoints differ from their own 18 Asked students to help each other understand ideas or concepts
III	<p>Establishing Rapport</p> <ul style="list-style-type: none"> 1 Displayed a personal interest in students and their learning 2 Found ways to help students answer their own questions 7 Explained the reasons for criticisms of students’ academic performance 20 Encourage student-faculty interactions outside of class (office visits, phone calls, e-mail, etc.)
IV	<p>Encouraging Student Involvement</p> <ul style="list-style-type: none"> 9 Encouraged students to use multiple resources (e.g. data banks, library holdings, outside experts) to improve understanding 11 Related course material to real life situations 14 Involved students’ in “hands-on” projects such as research, case studies, or “real-life” activities 19 Gave projects, tests, or assignments that required original or creative thinking
V	<p>Structuring Classroom Experience</p> <ul style="list-style-type: none"> 3 Scheduled course work (class activities, test, and projects) in ways which encouraged students’ to stay up-to-date in their work 6 Made it clear how each topic fit into the course 10 Explained course material clearly and concisely 12 Gave tests, projects, etc. that covered the most important points of the course 17 Provided timely and frequent feedback on tests, reports, projects, etc. to help students improve

Comparing Student Ratings in STEM and Non-STEM Classes

Researchers who have investigated differences in student ratings by academic discipline have found higher scores in the humanities and arts than in social sciences, which in turn are higher than in math and science (Braskamp & Ory, 1994; Cashin, 1990; Centra, 1993, 2009; Feldman, 1978; Hoyt & Lee, 2002b; Kember & Leung, 2011; Marsh & Dunkin, 1997; Sixbury & Cashin, 1995). One explanation for differences might be variance in quality of teaching. For example, in a study of IDEA ratings administered in classes from 2007-2011, Benton, Gross, and Brown (2012) found that instructors in STEM courses were less likely to encourage student involvement (e.g., hands-on projects, real-life situations), a teaching style associated with greater student learning. In addition, instructors in soft disciplines (e.g., political science, education) have been shown to exhibit a wider range of teaching behaviors than those in hard disciplines (e.g., engineering, chemistry) (Franklin & Theall, 1992). Moreover, instructors in the arts and humanities more frequently select objectives at the mid- and upper-levels of Bloom’s taxonomy of cognitive objectives (e.g., application, analysis) and tend to use active teaching methods. In contrast,

STEM faculty often rely more on lower-level objectives (e.g., knowledge, comprehension), and primarily use lecture.¹

But putting all the blame on instructors masks the other side of the story—the role of students who must certainly share some of the responsibility for lower ratings. Benton *et al.* (2012), for example, found that although students in STEM rated their courses as more difficult they did not report working any harder than students in non-STEM courses. So, although students acknowledge they find STEM domains more difficult, they apparently do not respond by exerting any more effort than they would in their other classes.

Another explanation for disciplinary differences in student ratings may be the sequential/hierarchical structure of content in STEM fields (Hativa, 2013b). Math and science, for example, have a structured knowledge sequence organized around well-established theories and principles. Students must rely upon what they have learned in prior courses to succeed in subsequent ones. Interestingly, however, instructors rate the adequacy of students' background and preparation similarly across STEM and non-STEM fields (Benton *et al.*, 2012). So, STEM faculty apparently consider their students no less prepared than do instructors in other disciplines.

Current Study

The current study compared IDEA SRI in STEM and non-STEM classes from 2009-2013. Only classes with a student response rate > 75% were included. Courses were offered under various formats (i.e., online, face to face, hybrid), and the ratings were based on both paper and online surveys. There were 171,306 STEM classes spread across the fields of science (82,200 classes), computer science (21,188 classes), engineering (12,444 classes), and math (55,474 classes). These were compared with 810,277 non-STEM classes. The highest degree awarded at the institutions included in the sample varied: associate (21%), baccalaureate (21%), masters (28%), and doctoral (30%). For the current analyses, only classes in which the principle type of student was first-year/sophomores (i.e., lower-level) were included.

Which objectives do instructors rate as Important or Essential in lower-level STEM and non-STEM courses?

Figure 9-1 shows the percent of STEM and non-STEM classes where the instructor selected each of 12 learning objectives as relevant (i.e., important or essential) to the course (see Table 9-1 for list of 12 learning objectives). Both STEM and non-STEM instructors tended to emphasize cognitive and application objectives (Objectives 1-4). Compared to non-STEM, STEM instructors placed relatively less emphasis on creativity and expressiveness (Objectives 6 and 8) and intellectual development (Objectives 7, 10, and 11). Perhaps most surprising is that only 44.7% of STEM instructors emphasized "Learning to analyze and critically evaluate ideas, arguments, and points of view" (Objective 11).

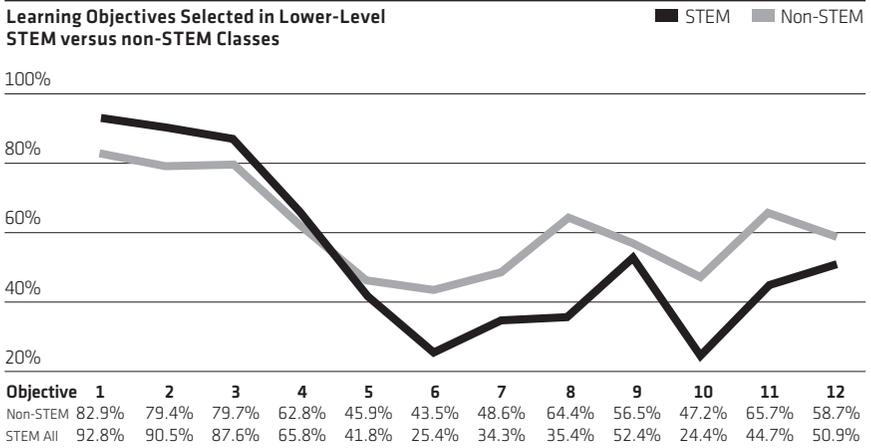


Figure 9-1

Percentage of total classes where instructor selected objective as “Essential” or “Important”.

When looking exclusively within STEM fields (Figure 9-2), again all disciplines, but especially math, stressed cognitive and application learning. In fact over 90% of math instructors stressed basic knowledge and its application, whereas only 50% or fewer emphasized any other outcomes. More than half of engineering instructors (61.5%) found team skills (Objective 5) important, whereas only about a fourth (24.1%) of math instructors did so. Finally, engineering and computer science placed more emphasis on information literacy (Objective 9) and life-long learning (Objective 12) than did other STEM disciplines.

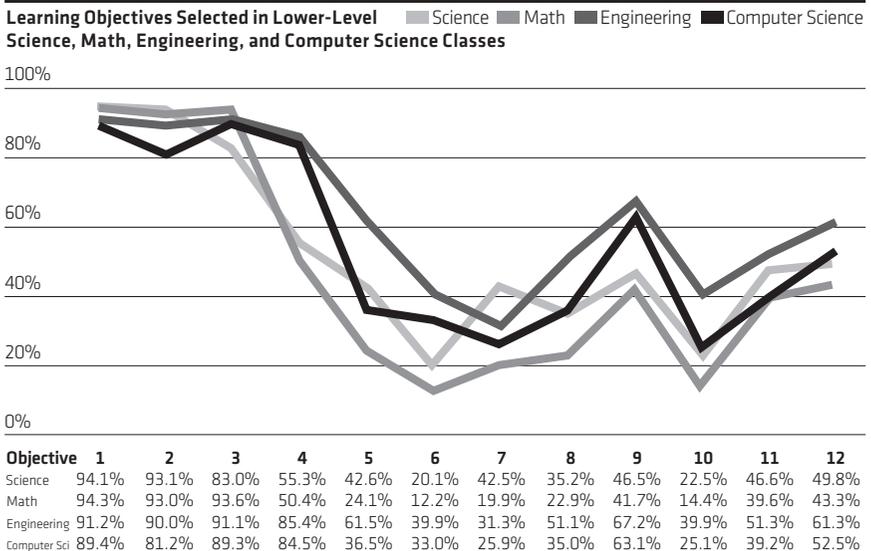


Figure 9-2

Percentage of total classes where instructor selected objective as “Essential” or “Important”.

How do student perceptions of their learning compare in lower-level STEM and non-STEM courses?

Figure 9-3 shows the percent of students in STEM and non-STEM classes that rated their progress on each of the 12 learning objectives as either exceptional or substantial. Only classes where the instructor rated the particular objective as either important or essential were included. Notably, comparing Figures 9-1 and 9-3 indicates that students tend to report the most progress on objectives the instructor emphasizes. This is consistent with previous research findings (Hoyt & Lee, 2002) and supports the validity of the IDEA system. Although students in STEM rated their progress lower on all objectives, they reported comparable progress to non-STEM students on cognitive, application, and problem-solving objectives (1-4). However, even when their instructors emphasized other types of learning (Objectives 5-12), students in STEM reported substantially less progress.

Examining STEM more closely, Figure 9-4 reveals some unique differences. Engineering students reported the most progress on all learning outcomes; whereas math students had lower ratings overall but especially in creativity and expressiveness.

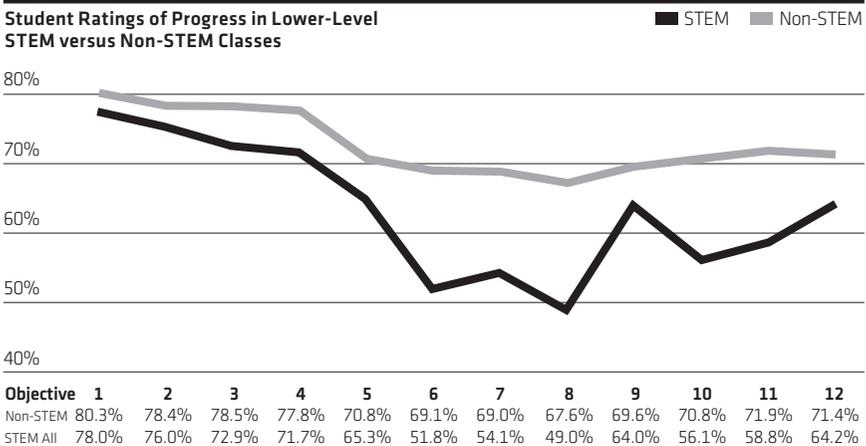


Figure 9-3

Percentage of total classes where instructor selected objective as “Essential” or “Important”.

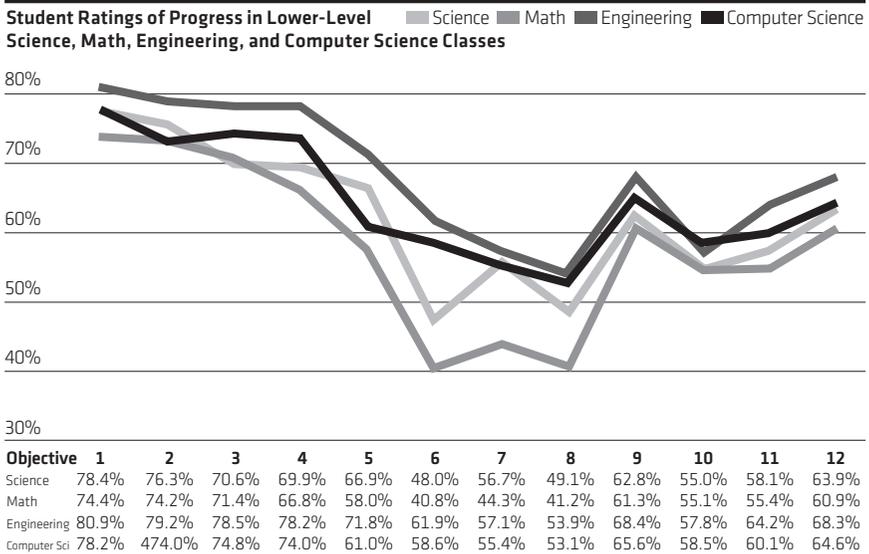


Figure 9-4

Percentage of total classes where instructor selected objective as “Essential” or “Important”.

Frequency of Teaching Styles Observed in STEM and non-STEM Courses

In the IDEA system, students report how frequently they observe each of 20 teaching methods associated with five teaching styles derived from factor analysis (Hoyt & Lee, 2002) and Chickering and Gamson’s (2007) principles of effective teaching (see Table 9-2). Each of these styles is positively but differentially correlated with student progress on relevant learning objectives (Hoyt & Lee, 2002). That is, certain styles are more highly correlated with progress on certain types of learning outcomes. “Stimulating student interest” and “Structuring the classroom experience,” for example, are the most important styles associated with student progress on cognitive learning objectives (Objectives 1 and 2) and in making applications of learning (Objectives 3 and 4) (Benton, Webster, Gross, & Pallett, 2010), which are the outcomes most frequently emphasized in STEM fields. Figure 9-5 shows that most STEM students observed these two styles almost always or frequently in their classes. So, there is congruence between the methods STEM instructors employ and the learning outcomes they emphasize.

Also depicted in Figure 9-5, far fewer STEM students reported their instructor fostered student collaboration and encouraged student involvement. That STEM instructors tended not to apply these styles in their classes may come as neither a surprise nor a concern, given such styles are not strongly associated with student progress on objectives STEM instructors typically emphasize (Objectives 1-4) (Benton *et al.*, 2010).

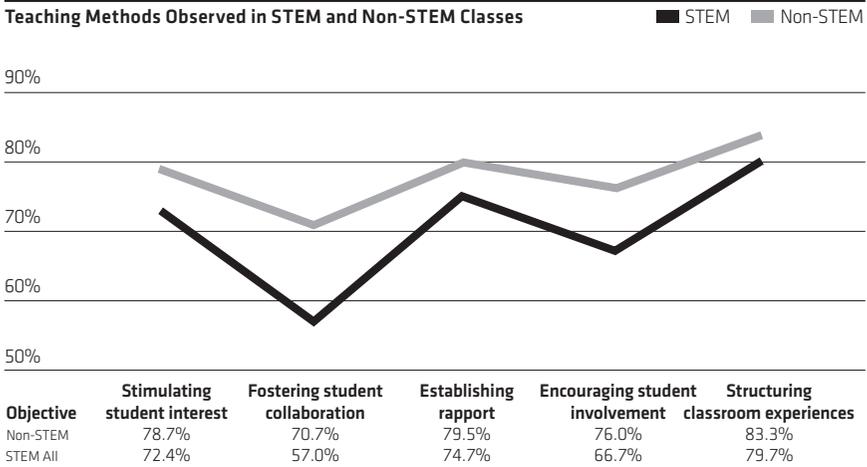


Figure 9-5

Percentage of total classes where teaching methods were observed.

Which teaching methods are important for student progress in lower-level STEM courses?

To answer this question, a larger dataset was analyzed that included IDEA student ratings from 488,415 classes across the years 2002 to 2011. Percentage breakdowns of STEM fields and types of institutions were similar to those reported previously. Bayesian Model Averaging (BMA) was used to provide estimated probabilities that each teaching method is associated with progress toward a given learning objective. The method by which BMA estimates probabilities helps to account for multicollinearity among the explanatory variables because it is based on computing multiple regression models. Since computing all possible models is unreasonable given the number of explanatory variables (*i.e.*, 20 teaching methods) and the vast size of the data set, the calculations were based on the best 100 models, where “best” was determined by Sawa’s Bayesian Information Criterion (SBC) values. The top few models accounted for most of the cumulative probability, and the rest accounted for vanishingly little of it.

The probability that a variable (*i.e.*, teaching method) belonged in the correct model was estimated as the sum of the model probabilities for all models in which that explanatory variable appeared. Hence, the more the explanatory variable appeared in the higher-probability models, the larger the probability that the respective variable belonged in the model. The regression parameters for a given explanatory variables were estimated by summing the regression parameters for that variable across all models in which it appeared, weighted by the estimated model probabilities. This is called the model-averaged estimate.

Table 9-3: Summary of BMA Modeling for First-Year/Sophomore Students, 2002-2011

Objective	Significant Teaching Methods Included in Full Models			
	Type 1 Students	R ²	Type 2 Students	R ²
1. Factual knowledge	8 (6,10,13)	.74	6,8 (10,12,13)	.75
2. Principles and theories	8 (6,10,13)	.75	8 (6,10,13)	.75
3. Applications	15 (2,4,10,18,19)	.76	(2,10,15,19)	.76
4. Professional skills, viewpoints	15 (2,6,7,10,14)	.75	(6,7,14,15)	.74
5. Team skills	5,14 (2,15,18)	.76	5,14,18	.74
6. Creative capacities	10,15,16,19	.75	14,15,19	.76
7. Broad liberal education	13,16 (8,9,10,15,19)	.76	10,15,16 (8,9,13,14)	.72
8. Communication skills	7,9,15,16,19	.73	9,16,19 (7,10,15)	.69
9. Find, use resources	9,15 (2,8,10,16,19)	.76	9 (2,15,16,19)	.75
10. Values development	15,16,19	.78	15,16	.77
11. Critical analysis	8,16,19 (10,15)	.75	16,19 (8, 10,15)	.74
12. Interest in learning	2,8,15 (10,13,16)	.77	15 (2,10,13,16,18,19)	.77

Note: Type 1 students are seeking to meet a general education or distribution requirement; Type 2 students are seeking to develop background needed for their intended specialization. Methods outside parentheses had model-averaged estimates > .10; those within parentheses had estimates > .05 and < .10.

Teaching Methods

1. Displayed personal interest in students and their learning
2. Found ways to help students answer their own questions
3. Scheduled course work (class activities, tests, projects) in ways which encouraged students to stay up-to-date in their work
4. Demonstrated the importance and significance of the subject matter
5. Formed “teams” or “discussion groups” to facilitate learning
6. Made it clear how each topic fit into the course
7. Explained the reasons for criticisms of students’ academic performance
8. Stimulated students to intellectual effort beyond that required by most courses
9. Encouraged students to use multiple resources (e.g., data banks, library holdings, outside experts) to improve understanding
10. Explained course material clearly and concisely
11. Related course material to real life situations
12. Gave tests, projects, etc. that covered the most important points of the course
13. Introduced stimulating ideas about the subject
14. Involved students in “hands on” projects such as research, case studies, or “real life” activities
15. Inspired students to set and achieve goals which really challenged them
16. Asked students to share ideas and experiences with others whose backgrounds and viewpoints differ from their own
17. Provided timely and frequent feedback on tests, reports, projects, etc. to help students learn
18. Asked students to help each other understand ideas or concepts
19. Gave projects, tests, or assignments that required original or creative thinking
20. Encouraged student-faculty interaction outside of class (office visits, phone calls, email, etc.)

For each model, BMA was conducted separately for two types of students based on instructor identification: first-year/sophomores seeking to meet a general education or distribution requirement (Type 1) and first-year/sophomores seeking to develop background in their intended specialization (Type 2). All data were analyzed at the class level. The models for IDEA Objectives 1 to 4, which are emphasized most frequently in STEM courses, were most relevant for the current study. Table 9-3 presents R-square values and summarizes the modeling across the 12 learning objectives.

Objective 1: Gaining factual knowledge. For both types of students, one teaching method stood out as most highly associated with student progress on this objective: stimulating students to intellectual effort beyond that required by most courses (#8). Making it clear how each topic fits into the course (#6) was also a strong predictor for Type 2 students. Their progress was also uniquely associated with instructors giving tests, projects, etc. that cover the most important points of the course (#12).

Objective 2: Learning fundamental principles, generalizations, or theories. The most important method associated with progress on this objective for both types of students is stimulating them to intellectual effort beyond that required by most courses.

Objective 3: Learning to apply course material. The greatest progress on this objective for Type 1 students is found in classes where the instructor frequently inspires students to set and achieve goals that really challenge them (#15); Type 2 students also benefit from this teaching method. Two teaching methods are uniquely tied to this group: demonstrating the importance and significance of the subject matter (#4) and asking students to help each other understand ideas or concepts (#18).

Objective 4: Developing skills, competencies, and points of view. The most important teaching method for Type 1 students inspiring them to set and achieve goals that really challenged them. Two other methods uniquely associated with progress by Type 1 students are finding ways to help them answer their own questions (#2) and explaining course material clearly and concisely (#10).

For information on helping students achieve IDEA Objectives 1 to 4 and on implementing each of the 20 teaching methods, see POD-IDEA Notes on Learning and POD-IDEA Notes on Instruction (<http://ideaedu.org/research-and-papers/pod-idea-center-notes-instruction>).

Conclusion

Students do not evaluate, administrators do. But students can provide valuable information about their perceptions of how much they believe they have learned and about what occurred in the course. This information can be helpful for making formative and summative decisions about teaching effectiveness when combined with other relevant indicators.

The current study of IDEA SRI showed several trends apparent across a wide variety of disciplines and institutions. First, instructors in STEM fields tend to emphasize cognitive background and applications of knowledge in their courses. With the exception of engineering, which also gives priority to information literacy and lifelong learning, most STEM instructors spend their time in the cognitive domain. Students, in turn, report their greatest progress on cognitive and application objectives. But they lag behind peers in non-STEM domains on other outcomes, even when their instructor finds those outcomes relevant to the course.

Congruence is found in the learning outcomes STEM instructors emphasize and the methods they employ most frequently in the classroom. STEM faculty generate interest by introducing stimulating ideas about the topic, inspiring students to achieve challenging goals, and demonstrating the importance and significance of the subject matter. They provide structure by making it clear how each topic fits into the course, explaining material clearly and concisely, and providing timely and helpful feedback. These teaching methods are highly correlated with student progress on cognitive and application objectives—the intended learning outcomes that receive the greatest emphasis in STEM courses.

BMA modeling reveals some differences between teaching methods helpful to lower-level STEM students seeking to meet general education or distribution requirements and those in the intended specialization. Although the findings are correlational, and therefore do not establish cause-effect, instructors may want to consider the reported relationships between teaching methods and student perceptions of their learning when they reflect on ways to enhance student achievement in STEM.

References

- Abrami, P. C., S. d'Apollonia, and S. Rosenfeld. (2007). The dimensionality of student ratings of instruction: An update on what we know, do not know, and need to do. In R. P. Perry and J. C. Smart (eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective* (pp. 385-445). Dordrecht, The Netherlands: Springer.
- Abrami, P. C., S. Rosenfeld, and H. Dedic. (2007). Commentary: The dimensionality of student ratings of instruction: What we know, and what we do not. In R. P. Perry and J. C. Smart (eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective* (pp. 385-446). Dordrecht, The Netherlands: Springer.
- Arreola, R. A. (2006). *Developing a Comprehensive Faculty Evaluation System (2nd edition)*. Bolton, MA: Anker Publishing.
- Benton, S. L., and W. E. Cashin (in press). Student ratings of instruction in college and university courses. In Michael B. Paulsen (ed.), *Higher Education: Handbook of Theory & Research* (Vol. 29). Dordrecht, The Netherlands: Springer.

- Benton, S. L., and W. E. Cashin. (2012). *Student ratings of teaching: A summary of research and literature*. IDEA Paper No. 50. Manhattan, KS: The IDEA Center.
- Benton, S. L., A. Gross, and R. Brown. (2012, October). *Which learning outcomes and teaching methods are instructors really emphasizing in STEM courses?* Presentation at American Association of Colleges and Universities Network for Academic Renewal, Kansas City, MO.
- Benton, S. L., R. Webster, A. B. Gross, W. Pallett. (2010). *Technical Report No. 15: An analysis of IDEA Student Ratings of Instruction in traditional versus online courses, 2002-2008 data*. Manhattan, KS: The IDEA Center.
- Braskamp, L. A., and J. C. Ory. (1994). *Assessing faculty work: Enhancing individual and institutional performance*. San Francisco: Jossey-Bass.
- Cashin, W. E. (1989). *Defining and evaluating college teaching*. IDEA Paper No. 21. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.
- Cashin, W. E. (1990). Students do rate different academic fields differently. In M. Theall and J. Franklin (eds.), *Student ratings of instruction: Issues for improving practice: New Directions for Teaching and Learning*, No. 43 (pp. 113-121). San Francisco: Jossey-Bass.
- Cashin, W. E. (2003). Evaluating college and university teaching: Reflections of a practitioner. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (pp. 531-593). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Centra, J. A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. San Francisco: Jossey-Bass.
- Chickering, A. W., and Z. F. Gamson. (1987). Seven principles for good practice in undergraduate education. *American Association of Higher Education Bulletin*, 39, 3-7.
- Davis, B. G. (2009). *Tools for teaching, (2nd ed.)*. San Francisco: Jossey-Bass.
- Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education*, 9, 199-242.
- Forsyth, D. R. (2003). *Professor's guide to teaching: Psychological principles and practices*. Washington, D.C.: American Psychological Association.
- Franklin, J., and M. Theall. (1992). *Disciplinary differences: Instructional goals and activities, measures of student performance, and student ratings of instruction*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Hativa, N. (2013a). *Student ratings of instruction: A practical approach to designing, operating, and reporting*. Oron Publications.
- Hativa, N. (2013b). *Student ratings of instruction: Recognizing effective teaching*. Oron Publications.
- Hoyt, D. P. and E. Lee. (2002a). *Technical Report No. 12: Basic data for the revised IDEA system*. Manhattan, KS: Kansas State University, The IDEA Center.

- Hoyt, D. P. and E. J. Lee. (2002b). *Technical Report No. 13: Disciplinary differences in student ratings*. Manhattan, KS: Kansas State University, The IDEA Center.
- Kember, D. and D. Y. P. Leung. (2011). Disciplinary differences in student ratings of teaching quality. *Research in Higher Education*, 52, 278-299.
- Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry and J. C. Smart (eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective* (pp. 319-383). Dordrecht, The Netherlands: Springer.
- Marsh, H. W. and M. J. Dunkin. (1997). Students' evaluations of university teaching: A multidimensional perspective. In R. P. Perry and J. C. Smart (eds.), *Effective teaching in higher education: Research and practice* (pp. 241-320). New York: Agathon Press.
- McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist*, 52, 1218-1225.
- Sixbury, G. R. and W. E. Cashin. (1995). *Technical Report No. 10: Comparative data by academic field*. Manhattan: Kansas State University, Center for Faculty Evaluation and Development.
- Svinicki, M. and W. J. McKeachie. (2011). *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers*. (13th ed.). Belmont, CA: Wadsworth

¹ Regardless of why disciplinary differences are consistently found in student ratings, proper use requires taking this into consideration when interpreting scores. Ideally, separate norms should be reported (as IDEA does in its course reports).

10

Enhancing Teaching and Learning: Potential and Reality

Marco Molinaro,
Assistant Vice Provost Undergraduate Education, University of California, Davis

Introduction

For many years the performance of students in STEM courses has been primarily focused on their individual outcomes and their ability or lack thereof. While there has been an interest in measuring the overall effectiveness of courses and instructors, the measure most commonly used has been the end of term classroom survey, an instrument which tends to place a great deal of emphasis on the enjoyment of a class or popularity of an instructor. Additionally, grading practices and the perceived value of gateway STEM courses to “weed” out unworthy students has led to populations of STEM graduates that are not representative of overall university or societal demographics. We believe it is time to acknowledge that student performance is not solely the responsibility of the students but intertwined in a system that involves the instructor, departmental norms and university support. Towards these ends, without improving the measurement of teaching practices and outcomes, there is likely to be little change. At the iAMSTEM Hub at UC Davis, we are engaged in developing new understanding and measures related to instruction and student outcomes. We strongly believe that without data that can speak to STEM instructors, department chairs and deans the status quo will prevail. While we fully acknowledge that data alone does not change culture, the lack of data allows very little to be visible and positive change becomes almost impossible to measure, let alone reward.

The iAMSTEM Hub

Who Are We and What We Do

The UC Davis iAMSTEM Hub is a university-wide STEM education effort working across relevant disciplines to maximize UCD graduates' capability and resilience through evidence-based actions. Our three primary goals are: 1) To catalyze necessary changes in institutional culture and policy while developing professional communities to support/sustain change; 2) To energize and implement necessary innovations in instruction, assessment, curriculum, and experiences; and 3) To build and share analytics tools and develop the architecture to measure and inform improvement of student outcomes and teaching practices. While originally focused on STEM undergraduate courses, the value of such an entity and its approaches is gaining broader acceptance with likely forays into economics and psychology gateway courses.

Where Are We Situated Within University

The Hub is explicitly placed at the level of reporting to the Provost and to Undergraduate Education to avoid any specific allegiance to a college or department while garnering the attention and resources that can make a difference at the level of the institution.

Measuring Instruction and Instructional Impact

What Happens in the Classroom

Student and Instructor Perceptions—WIDER: As part of an NSF grant we asked all instructors of the first two-year biology sequence and their students to answer a range of questions related to their approach and experiences in

Relative Importance of the Frequency of Biology Classroom Activities

4=Very Important 3=Important 2=Somewhat 1=Not

Both instructors and students (respectively) were asked: How important is it to you that the typical student does the following in your class? How important is it that you do the following to do well in your current biology class?	Show demonstrations, simulations, or video clips.	Actively engage in student-student or group discussions.	Connect learning to societal problems or issues.	Learn something that changes the way I understand an issue or concept.	Connect ideas from this course to prior experiences and knowledge.	Apply facts, theories or methods to practice problems or new situations.	Analyze an idea, experience, or line of reasoning in depth by examining its parts.	Form a new idea or understanding from various pieces of information.	Practice and/or explore multiple ways of answering questions and solving problems.	Practice using and interpreting multiple means of data representation (models, diagrams, graphs, etc.)
Instructor 1 BIS 2A Students	4 1.81	4 2.59	4 2.41	4 3.90	2 3.01	4 3.31	4 3.26	4 3.12	4 3.12	2 2.88
Instructor 2 BIS 2B Students	1 1.84	1 2.17	4 2.50	4 3.00	4 2.93	4 3.17	4 3.04	4 2.91	4 2.93	4 3.17
Instructor 3 BIS 2B Students	1 1.84	2 2.17	2 2.50	2 3.00	2 2.93	4 3.17	2 3.04	4 2.91	4 2.93	4 3.17
Instructor 4 BIS 2C Students	4 2.37	4 2.43	4 2.58	2 3.02	4 3.20	4 3.25	2 3.13	4 3.05	2 3.03	4 3.08

Table 10-1

Specific data from the initial Fall 2013 WIDER Survey highlighting differences between faculty and student perceptions.

the course. Initial analysis of results point to discrepancies in perceptions of amount of lecturing and levels of course interaction as well as differences in the levels of relevancy and importance of application in course content and delivery. See Table 10-1.

Student and Instructor Actions—COPUS: Self-reported perceptions of classroom activities tend to vary between students in a course and their instructors. In an effort to collect more objective data of classroom activities, we have adopted the COPUS instrument (CBE-LSE 2013 Winter;12(4):618-27. doi: 10.1187/cbe.13-08-0154) and have now created an online application to make two-minute observations of student and instructor action straightforward. Such observations undertaken multiple times in a quarter by multiple observers that are compared allow a measure of objectivity in determining the general pedagogical approaches utilized. See Figure 10-1 below and Figure 10-2 on the next page.

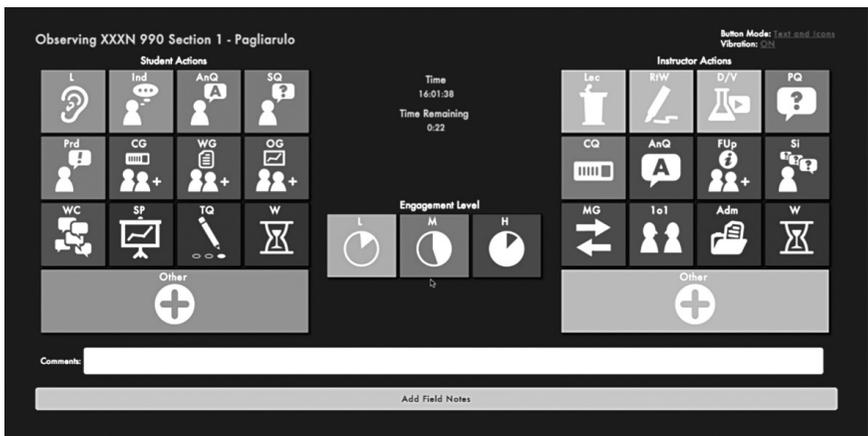


Figure 10-1
iAMSTEM online application that utilizes the COPUS protocol.

biology lecture with a traditional recitation type discussion versus an active student-centered discussion. As observed in many prior studies, traditional lecture tends to guide students towards a more novice view of the field after course completion. In the sections with the active discussions we were able to reduce the decline as well as increase the expert level response in a few dimensions (See Table 10-2). Since the presentation in early 2014 we have collected CLASS data for thousands more introductory biology and chemistry students and are in the process of analysis.

	Class A			Class B			Time X Group Interaction p-Value
	Pre	Post	Change	Pre	Post	Change	
Overall	63.4	60.9	-2.5	63.4	58.9	4.5	.075
Real World Connections	71.4	70.6	-0.8	73.7	68.0	-5.7	.010*
Problem-Solving Difficulty	49.1	47.1	-2.0	48.5	45.6	-2.9	.618
Employment	59.6	59.9	0.3	60.2	56.1	-4.1	.023*
Problem-Solving Effort	64.2	64.0	-0.2	66.1	61.2	-4.9	.012*
Conceptual Connections/Memorization	68.3	65.0	-3.3	69.5	64.0	-5.5	.203
Problem-Solving Strategies	67.0	69.6	2.6	68.1	65.9	-2.2	.042*
Reasoning	76.8	73.5	-3.3	78.3	73.1	-5.2	.384

Table 10-2

CLASS-BIO results for a late 2013 introductory biology class of approximately 600 students in each classroom A and B. Same instructor and material taught in both classes with Class A utilizing interactive discussion sections. Note: A statistically significant Time X Group Interaction indicates that the amount Class A and Class B changed from pre to post was significantly different.

Classroom Evaluations: Early analysis looking at course evaluations in chemistry has allowed us to determine, for a course taught by three different instructors, that individual student performance had very low correlation with overall perception of course or instructor ratings. Additionally, pre/post learning gains were not connected with instructor ratings. We have also learned that the departmental variation in questions asked, approach to measurement online and level of deidentification will make additional analyses into course evaluations very difficult at UC Davis.

Term by Term Variation

Grading Norms, Instructor and Student Variation: In introductory chemistry we have observed that controlling for incoming SAT characteristics there appear to be substantial variations in student course outcomes based on instructors. What we failed to realize at the time was much of the variation for most instructors is explained by differences in the student pools found

in specific quarter/time of day course offerings coupled with curved grading practices. Subsequent detailed analyses have found 0.5 GPA variations for equivalent students taking the introductory course purely based on their peers. Since the lower performers take the course “off-sequence” they tend to receive better grades for what may be reduced learning. This second part, related to the pre/post content measures, is currently being analyzed.

Instructor Impact on Follow-up Courses: Initial analyses show that there are interactions between instructors of the initial course of a sequence and subsequent courses that can result in up to 1.5 GPA variation. Further investigations of these effects have to be undertaken as a greater understanding of the consistency, or lack thereof, of testing and connectivity between courses is needed. We are investigating the best ways of representing subsequent course performance in a dashboard discussed at greater length in the next chapter.

Impact for Departments Over Time: Visualizing Student Flows

Tool Description: The ribbon flow tool helps to visualize amounts and paths from a starting point to an ending point. This tool can be used to visualize the *discipline/major in* of a starting cohort of students and the resulting *discipline/major out* after a specified amount of time. An example of the visualization shown on following page, Figure 10-3. The information presented can be interactively selected to emphasize a discipline or a set of majors on the starting and/or the ending side. At UC Davis, we are using this diagram to visualize components of a 120x120 matrix of in/out majors with the ability to define specific starting and ending criteria. The tool can work with any categorical data (including sub-categories such as discipline and major, in our case) with start point A, transitioning to a point B, and beyond. We also use it to visualize double majors to quickly understand the most popular options by discipline and/or major. All the “strands” can be queried by placing the cursor on top to get quick numbers and percentage information. The raw numerical data can be copied from a text box and directly pasted into any spreadsheet program.

Conclusions

Based on our initial work collecting a broader set of instructional metrics and presenting them to faculty we have observed rich instructional improvement discussions and an agreement to engage in instructional action/change. Because we are seeking a sustainable change/improvement in instruction and learning outcomes we need to create and enable a broadly usable approach that leads to a culture change where continuous improvement of instruction and learning is the new normal. Towards this end we have created a new set of data specific to instruction that when coupled with our historical and current individual student level data system enables us to account for course performance that is controlled for students in individual classes. As part of

Impact for Departments Over Time: Visualizing Student Flows

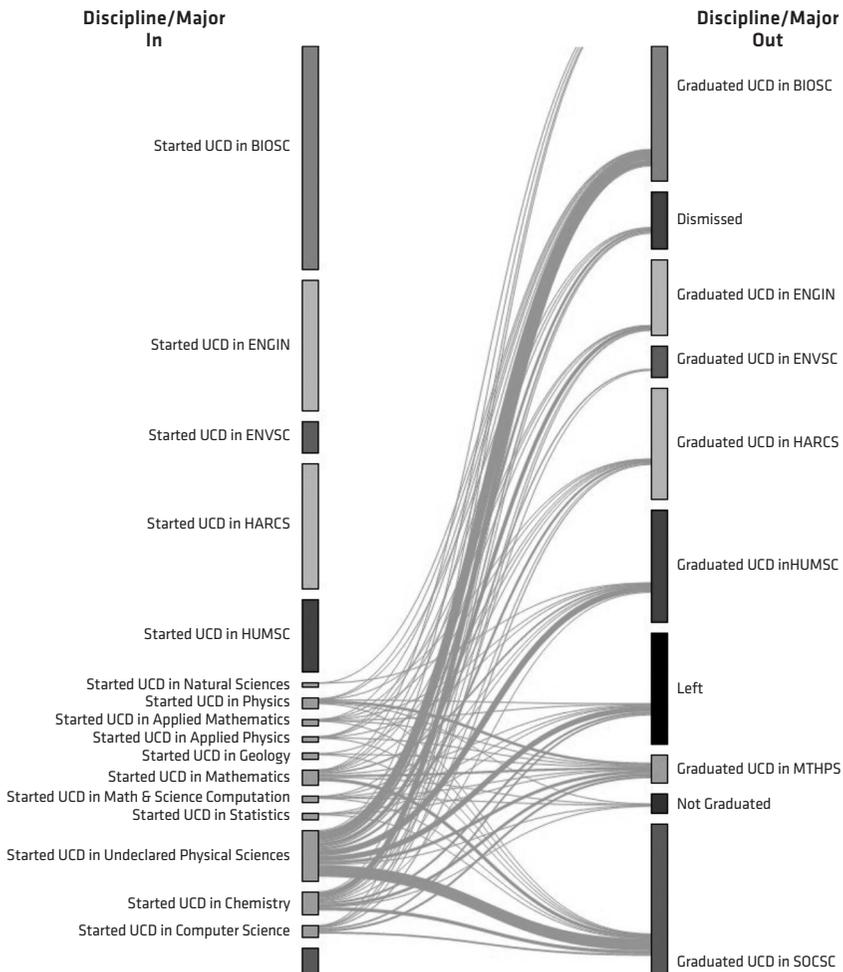


Figure 10-3
This is a visualization of the discipline/major in of a starting cohort of students and the resulting discipline/major out after a specified amount of time.

our longer term vision, we are starting to develop a system able to present each STEM gateway instructor with a “dashboard” that presents multiple performance measures on their individual courses, their historical trajectory/slope for a given course, how their measures fit within the departmental range for a given course and in a given course progression, and how their students do in the next course in a sequence. Indications thus far show that the instructional culture is changing, how far it will go and at what pace is yet to be determined.

11

In Search of Improved Student Learning: Strategies for Affirming We Are “On Track”

Jodi L. Wesemann,
Assistant Director for Educational Research, American Chemical Society

A migration towards improved student learning is underway in higher education. A series of parallel and iterative excursions are being taken through a complex and changing landscape. Some are undertaking solo expeditions, exploring what could be done in a specific course. Others are part of multi-year adventures with large groups of colleagues, some of whom may be from other departments or institutions. The activities of the Cottrell Scholars Collaborative, the AAU Undergraduate STEM Education Initiative, and other panelists at the Workshop on Effective Evaluation of Teaching and Learning are among the exciting adventures underway.

The answers to the questions of where to go, how to get there, and how to affirm that we are “on track” are unique and often interwoven. Several workshop panelists framed their remarks in the broad context, sharing stories about the learning outcomes and insights their departments, centers, institutions, or consortia are pursuing and how they are navigating the terrain. Other presentations focused specifically on direct and indirect ways people are assessing, as well as evaluating, efforts to improve student learning.

Just as there is no “right track” for educational reform, no one plan for assessing and evaluating teaching and learning will fit the various shifting landscapes in higher education. We need assessment approaches and tools that can indicate whether or not, and how efficiently and effectively, we are making progress towards our destination and help us decide if we should take a different route or even change our intended destination. If it is critical to reach a specific destination, we need appropriate standards or targets against which to benchmark and evaluation mechanisms for doing so. We need a

portfolio of activities—some that are near-term and easy to adapt to different contexts and others that are exploring new potential approaches. Multiple measures are needed to understand our impact and guide our next steps.

The need exists. A growing number of resources are available. Yet changing the ways we assess and evaluate teaching and learning remains difficult for several reasons.

- There is a lot at stake. Assessment and evaluation are integral parts of efforts to improve learning—whether solo or collective—and critical to their success. We are investing time and resources. We are taking risks. There is pressure to do it right.
- It is multifaceted. In addition to disciplinary perspectives, changing assessments and evaluations involves a range of educational, political, and often technical aspects.
- It requires change. Regardless of the nature or scope of the activities—whether customizing the end-of-course student questionnaires, adapting existing assessment tools, introducing course portfolios, or developing new tracking systems—we must do something different. It is not just implementing new assessments and evaluations. We must also respond to what we learn from them.

The workshop conveyed the need to understand the process of change and develop the skills to navigate and guide others through it. Presentations and discussions reflected three phases covered during “Leading Change,” a course in the ACS Leadership Development System, along with some general principles that can be applied regardless of the landscape or scope of efforts.

Creating a Sense of Urgency

The sense that we need to change the ways we assess and evaluate teaching and learning is increasing, thanks to growing pressures from students, faculty, staff, administrators, leaders of institutions, disciplinary societies, educational associations, funding agencies, governments, and the U.S. President. As Karen Bjorkman noted, many stakeholders want to know we are doing what we say we are going to do. Now we need to compel others to act, which they are more likely to do if they are personally concerned about the consequences of not changing. Noah Finkelstein raised the possibility of evaluations being imposed upon us—something that should prompt departments and institutions to take action. As Hunter Rawlings noted, our competitive spirits and concerns that others are getting ahead are also fostering a sense of urgency. All of the departments and institutions represented at the workshop are being encouraged to change by the increased visibility of questions, opportunities, and activities related to assessment and evaluation.

Describing a Winning Future

In addition to convincing our colleagues of the need to change the ways we assess and evaluate teaching and learning, we need to convey that it is possible and will be beneficial to do so. The experiences of others, shared via the literature, conferences, workshops, and reports, are invaluable sources of

inspiration and ideas. To be embraced, however, plans must be developed with input from those who will implement them. As Pratibha Varma-Nelson noted, there are no shortcuts.

It can be difficult, especially when we are eager to make changes, to take the time needed to consider and select from a range of options. Engaging as many of our colleagues as early as we can in a series of well-framed discussions about assessment and evaluation has three fundamental benefits. It can lead to a broader and richer range of options, get people on the same page, and foster the buy-in of those needed to move forward.

Putting It in Context: Since one size does not fit all, discussions about assessment and evaluation must consider our unique contexts—identifying specific needs and considering the status quo and how it is working, as well as what would happen if nothing was done differently. As Finkelstein noted, there are many grain sizes and timescales of needs that could be addressed. Initial discussions must define and focus the goals of near-term efforts, affirming our commitment to pursuing them.

Even when focused on well-defined goals, discussions about assessment and evaluation will generate a plethora of options—only some of which can be pursued with existing resources. The process of prioritizing these options also should be framed, starting with the criteria by which they will be considered. Is time of the essence? Are there resources that are short-lived? Do we need direct measures? What are the expectations of our administrators and funders? Clarifying these and other criteria before considering the options will help us make strategic and realistic choices about where to direct efforts and resources.

While remaining focused on clearly articulated needs, it is wise to be flexible. Our assessment approaches and tools, standards, and evaluation mechanisms should allow us to accommodate changes in the students, faculty, departments, and institutions involved, as well as to respond to inevitable shifts in the broader educational, scientific, technological, political, economic, and demographic landscapes.

Making It Meaningful: Our assessments and evaluations should be purposeful and authentic, addressing needs that students, faculty, institutions, or the broader community consider important. As Marco Molinaro emphasized, the results must add value. It is critical to help others understand how activities relate to their goals. Assessing initial activities and providing feedback can help improve the evaluation of performance, whether it is the grades for a course, tenure review, department review, project report, or accreditation review. When there are official institutional requirements, we need to look for ways to fulfill them that also inform individual and departmental efforts. As individuals, departments, centers, institutions, and a broader community, we must consider what we are going to do with the information collected.

Adjusting our tone and terminology can help frame efforts in a positive way. At the national level, conversations are shifting from teaching to learning and from content to outcomes/competencies. Other similar shifts were noted in the workshop. We are using assessments to identify strategies for improvement, rather than having evaluations simply indicate problems. While instructing students, we are also socializing them to academe and our disciplines. Rather than remediating students and correcting their misconceptions, we are empowering them.

We should do all we can to help others envision themselves as part of our plans—using a broad range of examples, defining terms, being receptive to different perspectives and approaches. Highlighting the value of input from all those involved helps make it feel like a team effort. When conversations get difficult, we can heed Bjorkman’s advice of refocusing on the students and their learning.

Taking Action

Efforts to move forward must be well-timed. If they are undertaken too soon, key players may not buy-in. If we wait too long, we lose momentum and other efforts take priority. By initially pursuing small but strategic actions and sharing the successes of those actions, interest in further actions will grow.

Building It Together: Faculty are interested in improving learning. Mary Ann Rankin shared what can occur when they are given permission to do things that they think need to be done. Overcoming the culture of the closed door, as Robin Wright noted, dispels the sense that colleagues do not care. Partnership for Undergraduate Life Sciences Education (PULSE) Fellows are guiding efforts called for in *Vision and Change in Undergraduate Biology Education: A Call to Action*. Learning communities are providing the safe spaces that Maura Borrego indicated we need. Including instructional and adjunct faculty, graduate students, and postdoctoral scholars prepares them to continue and build on current efforts.

As iAMSTEM and the peer observation projects demonstrate, input from a diverse group of stakeholders helps shape more robust plans that fit current and future contexts—while addressing biases and busting myths. Sharing available data, as was done at Washington University, generates opportunities for systemic collection. Engaging others as partners, as centers for teaching and learning and CUSTEMS are doing, facilitates buy in and establishes a shared vision that can sustain efforts over the long term.

The connection among small actions is important to note, whether we are focusing on individual classrooms and laboratories or comprehensive initiatives that span the educational enterprise. Aligning efforts at local and national levels increases their chances of success over the near- and long-terms. We should look for additional opportunities to leverage current and future investments being made in educational research and reform.

Developing Ongoing Processes: Some assessments and evaluations are readily implemented. Others require additional knowledge, discussions, and staging. Our efforts, just like those of our students, benefit from moving through the zones of proximal development that Chandralekha Singh noted. We are undergoing journeys, much like the discovery-based research experiences through which Scott Strobel guided his students.

As Molinaro noted, we don't just want to optimize current behavior. We must do all we can to facilitate the appropriate use of data and establish regular discussions of our assessments and evaluations, helping them remain relevant and grow. Input from students, colleagues, department reviews, advisory boards and steering committees, and national efforts allows us to adjust plans in ways that reflect the continuously changing landscapes, shifting external pressures, new opportunities, and the growing knowledge of teaching and learning. Such input also sustains the sense of urgency needed to continue driving change.

The migration towards improved student learning is gaining momentum. Assessments and evaluations promise to influence its direction and progress. As the workshop highlighted, there are roles for all of us to play, within classrooms and across higher education and disciplinary communities. Being aware of the concerns, the multiple facets, and the way changes occur helps us overcome barriers and maximize the impact of our efforts. Developing our skills prepares us to navigate the questions and tensions associated with education reform, balancing a range of immediate and systemic needs while being responsive and reflective. The ability to traverse boundaries, embracing cultures and languages of other disciplines, leads to changes that are well-informed and can have far-reaching impact. Collectively, we are helping each other develop the agency and support structures needed to pursue meaningful and evidence-based changes, affirm we are “on track,” and make informed investments that keep us moving in strategic directions.

The views expressed are those of the author and not necessarily those of the American Chemical Society.

Workshop Participants

Stephen Benton, *Senior Research Officer, IDEA Education*
Karen Bjorkman, *Dean and Distinguished University Professor of Astronomy, University of Toledo**
Maura Borrego, *Associate Professor of Mechanical Engineering, The University of Texas at Austin*
Myles Boylan, *Lead Program Director, Division of Undergraduate Education, NSF*
Stephen Bradforth, *Professor of Chemistry, University of Southern California**
Gail Burd, *Senior Vice Provost for Academic Affairs, University of Arizona*
Rocio Chavela, *Manager of Faculty Development, American Society for Engineering Education*
William R. Dichtel, *Associate Professor of Chemistry and Chemical Biology, Cornell University**
Peter Dorhout, *Dean of the College of Arts and Science, Kansas State University*
Susan Elrod, *Dean of the College of Science and Mathematics, California State University, Fresno*
Andrew Feig, *Associate Professor of Chemistry, Wayne State University**
Noah Finkelstein, *Professor of Physics, University of Colorado Boulder*
Catherine Fry, *Project Manager, AAC&U*
Howard Gobstein, *Executive Vice President, APLU*
Robert Hilborn, *Associate Executive Officer, American Association of Physics Teachers*
Jay Labov, *Senior Advisor for Education and Communication, National Academy of Sciences*
Adam Leibovich, *Professor of Physics, University of Pittsburgh**
Shirley Malcom, *Head, Education & Human Resources Programs, AAAS*
James Martin, *Professor of Chemistry, North Carolina State University**
Kathryn Miller, *Professor and Chair of Biology, Washington University in St. Louis*
Emily Miller, *Project Director, Association of American Universities**
Marco Molinaro, *Assistant Vice Provost for Undergraduate Education, UC Davis*
Lynne Molter, *Professor of Engineering, Swarthmore College*
Mary Ann Rankin, *Senior Vice President and Provost, University of Maryland, College Park*
Kacy Redd, *Director, Science and Mathematics Education Policy, APLU*
Zachary Schultz, *Assistant Professor of Chemistry, University of Notre Dame**
Susan Singer, *Director, Division of Undergraduate Education, NSF*
Chandralekha Singh, *Professor of Physics, University of Pittsburgh*
Linda Slakey, *AAU Senior Adviser, Association of American Universities*
Tobin Smith, *Vice President for Policy, Association of American Universities**
Matt Stephen, *STEM Project Assistant, Association of American Universities*
Scott Strobel, *Henry Ford II Professor of Molecular Biophysics and Biochemistry, Yale University*
Pratibha Varma-Nelson, *Professor of Chemistry and Executive Director,
Center for Teaching and Learning, Indiana University-Purdue University Indianapolis*
Jodi L. Wesemann, *Assistant Director for Educational Research, American Chemical Society*
Robin Wright, *Associate Dean, College of Biological Sciences, University of Minnesota*

* Member of the workshop planning committee.

About the Sponsors

The **Association of American Universities** (AAU) is a nonprofit association of 60 U.S. and two Canadian preeminent public and private research universities. Founded in 1900, AAU focuses on national and institutional issues that are important to research-intensive universities, including funding for research, research and education policy, and graduate and undergraduate education. AAU programs and projects address institutional issues facing its member universities, as well as government actions that affect these and other universities. The major activities of the association include federal government relations, policy studies, and public affairs. www.aau.edu



Inspired by our founder, Frederick Cottrell, **Research Corporation for Science Advancement** (RCSA) champions the best and brightest early career researchers in the physical sciences—astronomy, chemistry, physics and closely related fields. By providing highly competitive, significant research grants, RCSA encourages these young scientists to tackle globally significant problems that transcend traditional academic disciplines. Over the past century, RCSA, America's second-oldest foundation and the first devoted wholly to science, has supported the early career research of 40 Nobel laureates as well as thousands of academic and scientific leaders. Through its conferences and sponsored events, RCSA continues to advance America's scientific enterprise and the people who make it possible. Overseen by a Board of Directors composed of key leaders from finance, government and academic-based science, the Foundation and its awardees are helping to shape our nation's future in an increasingly competitive world. www.rcsa.org



RESEARCH CORPORATION
for **SCIENCE ADVANCEMENT**

A foundation dedicated to science since 1912.

Research Corporation for Science Advancement is a foundation that provides catalytic and opportunistic funding for innovative scientific research and the development of academic scientists who will have a lasting impact on science and society.

“The overarching objective of AAU’s Undergraduate STEM Education Initiative is to influence the culture of STEM departments at research universities so that faculty members use teaching practices proven by research to be effective in engaging students in STEM education and in helping them to learn. Successfully achieving this goal will require a major cultural shift whereby increasing emphasis is placed at our institutions on how faculty are supported, rewarded and recognized for their teaching in addition to their research. The partnership between AAU and the Cottrell Scholars Collaborative on the “Effective Evaluation of Teaching and Learning” and from which this workshop arose has proven tremendously valuable in paving the way for us to begin to see real progress in achieving the institutional change needed to improve undergraduate STEM education at major research universities.”

Tobin Smith, *Vice President for Policy, Association of American Universities*

“The Cottrell Scholar Collaborative was launched in 2011 with a main goal in mind—assisting Cottrell Scholars and other teacher-scholars nationwide to become outstanding science education practitioners. At the 2012 Cottrell Scholar Conference in Tucson, a group of Scholars led by Steve Bradforth, University of Southern California, Will Dichtel, Cornell University, and Adam Leibovich, University of Pittsburgh, identified the need for more effective evaluation measures to better promote excellence in teaching. The project started with the non-trivial premise of identifying practical strategies that are both easy to implement and disseminate broadly. For better leverage, a team of Cottrell Scholars was put together and they partnered with the AAU STEM initiative. They delivered! This project fits the spirit of the CS Collaborative perfectly well—teamwork leading to practical advice to help the best teacher-scholars to implement modern approaches that will, ultimately and collectively, transform the way science is taught in American colleges and universities.”

Silvia Ronco, *Program Director, Research Corporation for Science Advancement*