
Authors

Michael Dennin, Zachary D Schultz, Andrew Feig, Noah Finkelstein, Andrea Follmer Greenhoot, Michael Hildreth, Adam K Leibovich, James D Martin, Mark B Moldwin, Diane K O'Dowd, Lynmarie A Posey, Tobin L Smith, and Emily R Miller

Aligning Practice to Policies: Changing the Culture to Recognize and Reward Teaching at Research Universities

Michael Dennin,[†] Zachary D. Schultz,[‡] Andrew Feig,[§] Noah Finkelstein,^{||}
 Andrea Follmer Greenhoot,[¶] Michael Hildreth,[#] Adam K. Leibovich,[®]
 James D. Martin,^{**} Mark B. Moldwin,^{††} Diane K. O'Dowd,^{**} Lynmarie A. Posey,^{§§}
 Tobin L. Smith,^{|||} and Emily R. Miller^{||||*}

[†]Department of Physics and Astronomy and ^{††}Department of Developmental and Cell Biology, University of California, Irvine, Irvine, CA 92697; [‡]Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556; [§]Department of Chemistry, Graduate School, Wayne State University, Detroit, MI 48202; ^{||}Department of Physics and Center for STEM Learning, University of Colorado, Boulder, CO 80309; [¶]Department of Psychology and Center for Teaching Excellence, University of Kansas, Lawrence, KS 66045; [#]Research and Graduate Studies, College of Science, and Department of Physics, University of Notre Dame, Notre Dame, IN 46556; [®]Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15260; ^{**}Department of Chemistry, North Carolina State University, Raleigh, NC 27695; ^{††}Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI 48109; ^{§§}Department of Chemistry, Michigan State University, East Lansing, MI 48824; ^{|||}Association of American Universities, Washington, DC 20005

ABSTRACT

Recent calls for improvement in undergraduate education within STEM (science, technology, engineering, and mathematics) disciplines are hampered by the methods used to evaluate teaching effectiveness. Faculty members at research universities are commonly assessed and promoted mainly on the basis of research success. To improve the quality of undergraduate teaching across all disciplines, not only STEM fields, requires creating an environment wherein continuous improvement of teaching is valued, assessed, and rewarded at various stages of a faculty member's career. This requires consistent application of policies that reflect well-established best practices for evaluating teaching at the department, college, and university levels. Evidence shows most teaching evaluation practices do not reflect stated policies, even when the policies specifically espouse teaching as a value. Thus, alignment of practice to policy is a major barrier to establishing a culture in which teaching is valued. Situated in the context of current national efforts to improve undergraduate STEM education, including the Association of American Universities Undergraduate STEM Education Initiative, this essay discusses four guiding principles for aligning practice with stated priorities in formal policies: 1) enhancing the role of deans and chairs; 2) effectively using the hiring process; 3) improving communication; and 4) improving the understanding of teaching as a scholarly activity. In addition, three specific examples of efforts to improve the practice of evaluating teaching are presented as examples: 1) Three Bucket Model of merit review at the University of California, Irvine; (2) Evaluation of Teaching Rubric, University of Kansas; and (3) Teaching Quality Framework, University of Colorado, Boulder. These examples provide flexible criteria to holistically evaluate and improve the quality of teaching across the diverse institutions comprising modern higher education.

Research on how students learn and on learner-centered teaching practices is well documented in peer-reviewed scholarship (National Research Council, 2000; Doyle, 2008; Ambrose *et al.*, 2010; Brown *et al.*, 2014) and more recently highlighted in high-level policy reports and papers (Handelsman *et al.*, 2004; Singer *et al.*, 2012; Kober, 2015). Robust evidence shows that active-learning pedagogies are more effective than

C. Gary Reiness, *Monitoring Editor*

Submitted February 13, 2017; Revised August 24, 2017; Accepted August 25, 2017

CBE Life Sci Educ December 1, 2017 16:es5

DOI:10.1187/cbe.17-02-0032

*Address correspondence to: Emily R. Miller (Emily.miller@aau.edu).

© 2017 M. Dennin *et al.* CBE—Life Sciences Education © 2017 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

traditional lecture-based methods in helping students, including students from underrepresented backgrounds, learn more, persist in STEM (science, technology, engineering, and mathematics) fields, and experience higher rates of completing their undergraduate degrees (Lorenzo *et al.*, 2006; Haak *et al.*, 2011; Eddy and Hogan, 2014; Freeman *et al.*, 2014; Becker *et al.*, 2015; Trenshaw *et al.*, 2016). Grounded in this scholarship, many postsecondary institutions have launched institution-wide efforts to improve the quality and effectiveness of undergraduate teaching and learning. As commented by Susan Singer, former director of the Division of Undergraduate Education at the National Science Foundation, the landscape is filled with encouraging ways to transform undergraduate education (Singer, 2015).

Despite this movement toward developing and supporting systemic reform in undergraduate education, a majority of research university faculty members who teach undergraduate science and engineering classes remain inattentive to the shifting landscape. Student-centered, evidence-based teaching practices are not yet the norm in most undergraduate STEM education courses, and the desired magnitude of change in STEM pedagogy has not materialized (Henderson and Dancy, 2007; Dancy and Henderson, 2010; Dancy *et al.*, 2014; Anderson *et al.*, 2011; Singer *et al.*, 2012; Malcolm and Feder, 2016).

A critical factor impeding systemic improvements of undergraduate education, not only in STEM fields, is how teaching is considered in the rewards structure. Development of a coherent set of policies to guide the evaluation of a faculty member's work is a precondition for improving the merit and promotion processes. However, evidence shows a wide variation in commitment to and expectations for research, teaching, and service between and within research universities exists (Fairweather and Beach, 2002). Furthermore, stated policies that articulate the value of teaching have been insufficient to raise the attention paid to teaching. Merely espousing the value of teaching is not enough. Frequently, department, college, and university practices do not align with stated priorities in their formal policies (Fairweather, 2002, 2009; Huber, 2002).

Currently, faculty members at research universities tend to be assessed and promoted mainly on the basis of research success (Bradforth *et al.*, 2015). "Neglect of undergraduate education had been built into the postwar university, in which faculty members were rewarded for their research output, graduate student Ph.D. production, and the procurement of external research support, but not for time devoted to undergraduate education" (Lowen, 1997, p. 224). This reality is frequently reinforced by a lack of support and feedback about teaching (Gormally *et al.*, 2014). Furthermore, teaching effectiveness is overwhelmingly assessed using student evaluation surveys completed at the end of each course, despite evidence that these evaluations rarely measure teaching effectiveness (Clayson, 2009; Boring *et al.*, 2016), contain known biases (Centra and Gaubatz, 2000), promote the status quo, and in some cases reward poor teaching (Braga *et al.*, 2014). There is also increasing evidence that unintended biases of students influence course evaluations (MacNeill *et al.*, 2015). However, the ease with which these student evaluation surveys are administered and used in the promotion and tenure process has resulted in a long-standing practice that presents a barrier to innovation and scholarly teaching.

Providing faculty members with support for improved teaching, aligning incentives with the expectation of quality teaching, using metrics that accurately reflect teaching effectiveness, and developing transparent evaluation practices that are not unduly biased are necessary for systemic improvement of undergraduate education. Enabling effective evaluation of teaching will require the development of practical frameworks that are scholarly, accessible, efficient, and aligned with local cultures so as not to preclude their use by most institutions. Such frameworks will provide the greatest probability that teaching and its evaluation will be taken seriously in the academy (Wieman, 2015).

This essay discusses the collaboration between the Association of American Universities¹ (AAU) and the Cottrell Scholars² funded by Research Corporation for Science Advancement (RCSA) to address this critical barrier to improve the quality of undergraduate STEM education. While our efforts have focused primarily on undergraduate STEM education, many of the recommended practices would serve to improve undergraduate education generally.

An initial AAU and Cottrell Scholar collaborative project (2012–2015) focused on understanding the landscape of established and emergent means to more accurately evaluate and assess teaching effectiveness. Building from this work, a second collaborative project (2015–2017) aimed to develop practical guidelines to recognize and reward contributions to teaching at research universities at the department, college, and university levels. A starting point for this project was to assess the current espoused importance of teaching at research universities by examining published promotion and tenure policies at research-intensive institutions. This information was combined with an analysis from a 2014 survey administered to instructional staff on the importance of teaching at research universities as part of the AAU Undergraduate STEM Education Initiative. These results formed the basis for a workshop sponsored by the AAU and RCSA held in May of 2016 that brought together leading higher education scholars and practitioners and research-active faculty members to develop specific recommendations and guidance to value, assess, and reward effective teaching.

The following essay reports on the gap between policies and practices within an institution and offers strategies intended to provide guidance on how institutions can more effectively align their practices for valuing teaching with the stated priorities in their formal policies. The essay concludes with profiles of three institutional examples drawing upon such strategies to assess and reward contributions to teaching.

¹Founded in 1900, the AAU comprises 62 distinguished institutions in the United States and Canada that continually advance society through education, research, and discovery. Our U.S. member universities earn the majority of competitively awarded federal funding for academic research, are improving human life and well-being through research, and are educating tomorrow's visionary leaders and global citizens. AAU members collectively help shape policy for higher education, science, and innovation; promote best practices in undergraduate and graduate education; and strengthen the contributions of research universities to society.

²The Cottrell Scholar program develops outstanding teacher-scholars in chemistry, physics, and astronomy who are recognized by their scientific communities for the quality and innovation of their research programs and their academic leadership skills.

THE GAP BETWEEN POLICY AND PRACTICE

An analysis of 51 institutions' university-level promotion and tenure policies shows that many contain language valuing teaching in addition to research (see list of universities in the Supplemental Materials). Forty-one of these policies give some form of guidelines as to how teaching should be considered. Out of the 41 institutions that provide guidelines, 36 required at least one form of evidence, 36 recommend or require student evaluations to be used, and 26 recommend or require peer classroom observation.

The AAU, as part of its Undergraduate STEM Education Initiative, collected statements on the evaluation of teaching from 32 department chairs at eight universities. Across all institutions and departments there was a strong assertion that teaching is highly valued. Furthermore, all departments make use of student evaluations at the end of courses and provide an annual award for excellence in teaching. However, it was impossible to discern for 19 of 32, or 59%, of the statements submitted whether attention to student learning outcomes or evidence-based pedagogy was either required or recognized.

Additionally, the AAU collected information about the value placed on teaching and the quality of the evidence used to assess effective teaching in merit and promotion processes from approximately 1000 instructional staff. Respondents³ were asked to indicate the extent to which they agreed with a series of statements about the value placed on teaching by their departments, colleges, and schools, as shown in Table 1. This information was collected to provide a baseline of the overall culture toward teaching at these various levels as part of the AAU Undergraduate STEM Education Initiative pilot project sites. Respondents agreed that both their department and campus administrations at their universities recognize the importance of teaching and are supportive of faculty members improving and changing their teaching practices (3.20 ± 0.74 and 3.02 ± 0.75 , respectively). However, when asked whether faculty members in their departments believe that ongoing improvement in teaching is part of their job duties, the level of agreement drops slightly (2.90 ± 0.74). Also, when asked to give their opinion whether effective teaching plays a meaningful role in the annual review and salary processes within their colleges and within the promotion and tenure processes at their institutions, the mean responses were in the middle between agree and disagree (2.50 ± 0.87 and 2.54 ± 0.86 , respectively). These results suggest some disconnection between what is publicly supported within colleges and universities and what actually happens in day-to-day processes.

Furthermore, when respondents were asked to provide their opinions about the quality of the evidence for effective teaching used by their colleges in annual review and salary processes and

TABLE 1. Overall means for survey statements by faculty members about importance and recognition of teaching (1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree)

Statement	Mean	SD	Valid N
My departmental administration recognizes the importance of teaching and is supportive of faculty improving and changing teaching practices.	3.20	0.74	964
Campus administration at my university recognizes the importance of teaching and is supportive of faculty improving and changing teaching practices.	3.02	0.75	960
Instructors in my department believe that ongoing improvement in teaching is part of their jobs.	2.90	0.74	962
In my opinion, effective teaching plays a meaningful role in the annual review and salary processes in my college.	2.50	0.87	950
In my opinion, effective teaching plays a meaningful role in the promotion and tenure processes at my institution.	2.54	0.86	950

in the promotion and tenure processes at their institutions, those choosing “don't know” or not answering increased to slightly more than 40% (see Table 2). Of those who chose to respond, in both cases, one-third noted the teaching evidence was of “low quality” and half cited “medium quality” evidence of effective teaching. This reinforces findings previously demonstrated by Wieman (2015).

RECOMMENDATIONS TO VALUE, ASSESS, AND REWARD CONTRIBUTIONS TO TEACHING AT RESEARCH UNIVERSITIES

Larger long-term improvement to undergraduate STEM education will evolve from an environment of continuous improvement of teaching coupled with an altering of the practice of how contributions to teaching are recognized and rewarded at research institutions. Interpretation and enactment of written policies relating to the evaluation of teaching for purposes of merit and promotion are where true institutional values lie (Fairweather, 2002).

Fostering a university culture that values high-quality and continuous improvement of teaching as much as performing high-quality research requires establishing teaching as a public and collaborative university activity, as well as an integral aspect of the individual faculty member's *scholarship*. To do this, it is critical to identify the criteria and relevant roles of the faculty member, program, department, college, and institution for evaluating an individual faculty member's work that fits both the local context (program/department/college) and the larger institutional mission. Ultimately, the goal is to allow local variation in a manner that both preserves the academic freedom of faculty in the classroom while supporting the university's collective responsibilities for undergraduate education.

³A total of 2971 instructional staff received the AAU faculty survey across the eight project-site institutions. More than 1000 (1093) submitted at least a partially completed survey, resulting in an overall response rate of 36.8%; individual institutional response rates ranged from 21.6% to 69.4%. A majority of respondents (542, or 49.6%) were either associate professors or professors with tenure. Twelve percent were tenure-track professors who did not yet have tenure at the time they were surveyed. More than a quarter of respondents were graduate students (26%), and the final 12.5% were instructor/lecturers, nontenured faculty, no response, or other instructional staff. Responses from private institutions comprised 36% of the total, with 64% from public institutions. The Supplemental Materials include the survey instrument and the complete project site baseline data report.

TABLE 2. Responses to quality of evidence of effective teaching

Your feedback regarding the quality of the evidence for teaching used in the following circumstances:	Low quality		Medium quality		High quality		Total	No response or don't know
	N	%	N	%	N	%	N	N
	By your college in the annual review and salary process	224	34.4	331	50.8	97	14.9	652
By your institution in the promotion and tenure process	212	33.2	325	50.9	101	15.8	638	455

Deans and Department Chairs Play a Critical Role

As institutional leaders, deans and department chairs can reinforce an expectation that faculty members understand teaching not as an isolated activity, but as integrated into their roles as scholars, as members of the university, and as members of their own disciplinary-based community. This requires department chairs and deans to encourage faculty members to think critically about their teaching and develop a continuous improvement mind-set about their teaching within their disciplines and in the context of the educational responsibilities of their departments. Beginning this conversation during the hiring process (e.g., through clear language in the job announcement and application package materials) articulates the importance of teaching. Furthermore, assessing a candidate's attitudes about teaching and advising can be achieved by including questions about teaching and advising in addition to research in the on-campus interview. Some schools also schedule a teaching demonstration (such as a mini-class) as part of the interview process, going beyond simply discussing teaching with the candidate. This approach demands that universities, colleges, and departments must desire to hire outstanding scholars who participate in the dissemination of the knowledge that they create and view teaching as an essential element of scholarship.

Emphasize the Importance of Teaching at the Onset of Hiring

For new hires, a department could provide teaching professional development funds as part of start-up packages, require a professional development plan for teaching, support participation in faculty learning communities, or intentionally support faculty mentoring by pairing expert teachers with those interested in improving their teaching and provide course-load credit for both faculty members.

Communicate Criteria and Expectations on How Contributions to Teaching Will Be Evaluated and Recognized

Faculty members should be provided with mechanisms to document and evaluate teaching innovations and improvements necessary to satisfy these criteria and expectations. Additionally, data from such documentation should feed into reward systems. Three practices are essential to this recommendation. 1) Empower departments to establish an agreed-upon set of metrics that go beyond student satisfaction surveys for each faculty member. A broader array of materials could include: development/revision of learning goals and content in course syllabus, incorporation of new pedagogical

practices into courses, documented achievement of student learning outcomes or changes in classroom culture, involvement in teaching service or scholarship, or shifts of assessment from factual recall to providing evidence of how students use their knowledge. The primary purpose of these strategies is to encourage faculty members to be reflective about their teaching practices. 2) Make sure that metrics are efficient, that is, they are not so labor-intensive as to preclude their use by most faculty members. 3) Ensure that promotion and tenure committees at both the departmental and institutional levels are educated with respect to best practices about how to effectively review the materials submitted by faculty members.

Establish a Culture Consistent across Departments, Colleges, and the University That Recognizes the Scholarly Activity of Teaching

Fundamentally, the values of a university and a department can be discerned from the activities they promote and reward. The above recommendations are aimed at establishing a culture consistent across departments, colleges, and the university that recognizes the scholarly activity associated with the time and effort to maintain and improve education. Achievement of this goal will require a holistic approach to value, support, assess, and reward teaching at multiple institutional levels. Universities and colleges can signal a commitment to quality educational practices for all by providing resources to support faculty members improving large introductory STEM courses. A commitment by the department and university to use clearly articulated empirical evidence for rewarding teaching, both in the promotion and tenure process and for teaching awards, provides validation for the importance of effective undergraduate education. Fundraising around curricular programs can bring exposure and reward to faculty members invested in student learning. Efforts to address the perceived divide between tenure-stream "research" faculty and instructional faculty, who often play a significant role in the large introductory courses, could further support the university's educational mission. Opportunities to discuss and present scholarly activities around teaching provide public recognition that can be emphasized by the visible support of key institutional leaders, such as deans, chairs, and other academic administrators. Furthermore, increasing awareness within the university about existing efforts and related scholarship to improve student learning and teaching effectiveness on campus has the potential to better articulate how the institution's educational objectives relate to the research mission of the university.

THREE EXAMPLES OF INSTITUTIONAL INITIATIVES TO ASSESS AND REWARD TEACHING

Promotion Process at the University of California, Irvine: Moving to a Three-Bucket System

At most universities, accomplishment in three areas—research, teaching, and service—is evaluated to inform merit and promotion decisions. Overall accomplishment that has both quantity and impact components can be represented by a single bucket (Figure 1A). The level that must be achieved for promotion varies by university and discipline, but is generally agreed upon locally and is represented by the dashed line in Figure 1A. Getting over the line results in promotion. But this graphic illustrates the common perception, particularly at R1 universities, that the full line can be determined almost completely by accomplishments in research.

When all faculty are compared in this one-bucket system, those who do more teaching and service rarely benefit in terms of merit and promotion, because getting to the dashed line is what is needed. One solution is to move to a three-bucket system, in which a level of accomplishment that has both quantity and impact components is required in each of three buckets (Figure 1B). If the faculty member does not reach the required level in all three buckets, merit-based salary increases are not awarded or promotion/tenure is denied. In this system, one cannot simply fill the research bucket so full that empty teaching and service buckets are acceptable.

A) One-Bucket System



B) Three-Bucket System

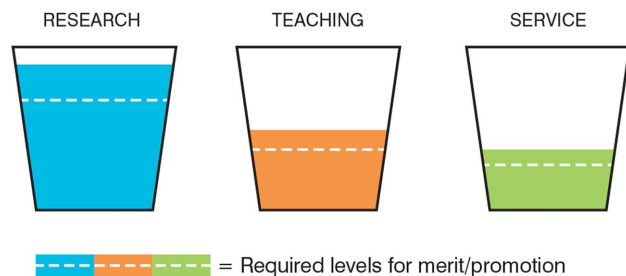


FIGURE 1. Moving from a one- to three-bucket system. (A) In the one-bucket system, the arrow on left indicates the level of accomplishment, determined by quantity and impact components, required for promotion. Sufficient accomplishment in research is often enough to reach this level. (B) Using a three-bucket system requires accomplishment not only in research but in teaching and service as well, and the shading indicates that accomplishment expected might vary depending on department, school/unit, or even at different times in one's career.

The University of California (UC), Irvine, has not yet made a complete transition from a one-bucket to a three-bucket system, but is making steady changes in this direction. For example, as a member of the UC system, UC Irvine has a merit and promotion system that governs advancement through the ranks with associated salary increases on a regular schedule (www.ucop.edu/academic-personnel-programs/_files/apm/apm-210.pdf). Advancements can be accelerated in time to reward the most outstanding faculty. On the UC Irvine campus, accelerations have typically required demonstration of research accomplishments at a significantly higher rate and of similar or greater impact than expected for a regular action. Since 2014, accelerations have required evidence of excellence above that expected for normal actions, not only in research but also in teaching and/or service.

What is put into the buckets also matters. While published UC policy indicates that at least two types of evidence should support evaluation of teaching (see the Supplemental Material or visit www.ucop.edu/academic-personnel-programs/_files/apm/apm-210.pdf, p. 5), in practice, student evaluations are often the only evidence used. For the 2016 review cycle, UC Irvine has required individuals to upload at least one additional type of evidence to evaluate teaching (e.g., reflective teaching self-statement, syllabus, peer evaluation, or measure of student achievement). This change is a first step toward conducting a more thorough evaluation of the contributions to teaching. It also broadens the discussion of teaching by everyone involved in the review process and thus has the potential to increase awareness of the innovative and effective teaching practices taking place on campus.

University of Kansas Department Evaluation of Faculty Teaching Rubric

The Center for Teaching Excellence at the University of Kansas (KU) recently developed a rubric (Figure 2) for department-level evaluation of faculty teaching. The university requires that evaluation of faculty teaching for promotion and tenure and progress toward tenure includes information from the instructor, students, and peers. However, the quality of the information collected is highly variable, and reviewers often struggle to integrate and make sense of information from the three sources. In practice, many evaluations prioritize a narrow dimension of teaching activity (the behavior of the instructor in the classroom) and a limited source of evidence (student evaluations). Providing a rubric to structure the evaluation of faculty members' teaching increases the visibility of all dimensions of teaching, clarifies faculty teaching expectations, enables quick identification of strengths and areas for improvement, and brings consistency across evaluations and over time.

The goal of the rubric is to help department committees integrate information from the faculty members being evaluated, their peers, and their students to create a more holistic view of a faculty member's teaching contributions. Drawing on the peer review of teaching literature (e.g., Hutchings, 1995, 1996; Glassick *et al.*, 1997; Bernstein and Huber, 2006; Bernstein, 2008; Lyde *et al.*, 2016), the rubric identifies seven dimensions of teaching practice that address contributions to both individual courses and the department's curriculum. For each category, the rubric provides both guiding questions and defined expectations. The rubric can also be used to guide a constructive peer-review process, reflection, and iterative improvement.

		Exceeds Expectations: 4-5		
		Meets Expectations: 3		
		Below Expectations: 1-2		
Expectation levels align with KU's promotion and tenure rating scale.		(1) Poor: Consistently at this level	(2) Marginal: Some teaching at this level	
Goals, content, and alignment	<ul style="list-style-type: none"> • What are students expected to learn from the courses taught? • Are course goals appropriately challenging? • Is content aligned with the curriculum? 	<ul style="list-style-type: none"> • Course goals are unclear, inappropriate, or marginally related to curriculum • Content and materials are outdated or unsuitable for students in the courses • Range of topics is too narrow or too broad • Content is not clearly aligned with curriculum or institutional expectations 	<ul style="list-style-type: none"> • Course goals are articulated and appropriate for curriculum • Content is current and appropriate for topic, students, and curriculum • Course topics include an appropriate range • Standard, intellectually sound materials 	<ul style="list-style-type: none"> • Course goals are well articulated, high quality, and clearly connected to program or curricular goals • Content is challenging and innovative or related to current issues and developments in field • Topics are of appropriate range and depth, with integration across topics • High quality materials, well-aligned with course goals
Teaching practices	<ul style="list-style-type: none"> • How is in-class and out-of-class time used? • What assignments, assessments and learning activities are implemented to help students learn? 	<ul style="list-style-type: none"> • Teaching practices are not sufficiently planned or organized, or are poorly implemented • Practices are not well executed; little development in methods despite evidence of need • Students lack opportunities to practice the skills embedded in course goals • Student engagement is variable 	<ul style="list-style-type: none"> • Teaching practices are well planned and organized • Standard course practices carried out; follows conventions within discipline and institution • Students have some opportunities to practice skills embedded in course goals • Students consistently engaged 	<ul style="list-style-type: none"> • Activities are well planned, integrated, and reflect commitment to providing meaningful assignments and assessments • Uses effective, high-impact or innovative methods to improve understanding • In- and out-of-class activities provide opportunities for practice and feedback on important skills and concepts • Students show high levels of engagement
Achievement of learning outcomes	<ul style="list-style-type: none"> • What impact do these courses have on learners? • What evidence shows the level of student understanding? 	<ul style="list-style-type: none"> • Insufficient attention to student learning—quality of student learning is not described or analyzed with clear standards • Evidence of poor student learning: low level of skill/understanding is required or achieved without clear attempts to improve 	<ul style="list-style-type: none"> • Clear standards for evaluating the quality of student understanding • Typical student achievement for courses at these levels 	<ul style="list-style-type: none"> • Standards for evaluating student understanding are connected to program or curriculum expectations, or use authentic assessments • Efforts to support learning in all students • Quality of learning supports success in other contexts (e.g., subsequent courses or non-classroom venues), or is increasing over successive offerings
Classroom climate and student perceptions	<ul style="list-style-type: none"> • What are the students' views of their learning experience? • How has student feedback informed the faculty member's teaching? 	<ul style="list-style-type: none"> • Classroom climate does not promote civility or discourages student motivation and engagement • Consistently negative student reports of teacher accessibility, interaction skills • Poor sense of learning among students • Little attempt to address concerns voiced by students 	<ul style="list-style-type: none"> • Classroom climate promotes civility • No consistently negative student ratings of teacher accessibility, interaction skills • Most students indicate progress with their learning • Instructor articulates some lessons learned through student feedback 	<ul style="list-style-type: none"> • Evidence that classroom climate is respectful, cooperative, and encourages motivation and engagement • Student feedback on teacher accessibility, interaction skills is generally positive • Students perceive that they are learning important skills or knowledge • Instructor is responsive to student feedback in short- and long-term
Reflection and iterative growth	<ul style="list-style-type: none"> • How has the faculty member's teaching changed over time? • How has this been informed by evidence of student learning? 	<ul style="list-style-type: none"> • No indication of having reflected upon or learned from prior teaching or feedback 	<ul style="list-style-type: none"> • Continued competent teaching, possibly with minor reflection based on input from peers and/or students • Articulates some lessons learned from prior teaching and feedback 	<ul style="list-style-type: none"> • Regularly makes adjustments to teaching based on reflections on student learning, within or across semesters • Examines student performance following adjustments • Reports improved student achievement of learning goals based on past course modifications
Mentoring and advising	<ul style="list-style-type: none"> • How effectively has the faculty member worked individually with UG or graduate students? 	<ul style="list-style-type: none"> • No indication of effective mentoring or advising students (but expected in department) 	<ul style="list-style-type: none"> • Some evidence of effective advising and mentoring (define as appropriate for discipline) 	<ul style="list-style-type: none"> • Evidence of exceptional quality and time commitment to advising and mentoring (define as appropriate for discipline)
Involvement in teaching service, scholarship, or community	<ul style="list-style-type: none"> • In what ways has the instructor contributed to the broader teaching community, both on and off campus? 	<ul style="list-style-type: none"> • No interaction with broader community about teaching including involvement with teaching-related committees • No evidence of keeping up with reports on effective teaching • Practices and results of teaching are not shared with others • Actions have negative impact on teaching culture in department or institution 	<ul style="list-style-type: none"> • Some involvement in teaching-related committees, or engagement with peers on teaching (e.g., teaching-related presentations or workshops) • Participates in department-level curriculum decisions 	<ul style="list-style-type: none"> • Regular involvement in teaching-related committees, engagement with peers on teaching (e.g., teaching-related presentations or workshops) • Occasional (or more) local or external presentations or publications to share practices or results of teaching • Contributes to department or university curricular planning or assessment • Advanced-Scholarly publications or grant applications related to teaching

FIGURE 2. Rubric for department evaluation of faculty teaching.

To ensure applicability across disciplines, the rubric does not weigh or place focus on any particular element or require a particular type of evidence to be used. Departments are encouraged to modify the rubric and use it to build consensus about the dimensions, the questions, and the criteria. The implementation strategy included discussions with department chairs and the KU Center for Teaching Excellence department ambassadors in advance of its release in order to increase the probability of broad buy-in. The rubric was piloted during the 2016–2017 academic year as a guide for peer review of teaching, promotion and tenure, and third-year reviews.

University of Colorado Teaching Quality Framework

The AAU-sponsored undergraduate STEM initiative has helped to support the development of a new Teaching Quality Framework (TQF) at the University of Colorado. The framework draws upon organizational change literature and cites the University of Colorado, Boulder (CU Boulder), already existing guidelines—“Dossiers for comprehensive review, tenure, or promotion must include multiple measures of teaching”—to create a framework for assessing and promoting teaching quality (Finkelstein *et al.*, 2015). The goal of the TQF is to create a common campus-wide framework for using scholarly measures of teaching effectiveness that is disciplinary specific and provides faculty members with feedback in order to support improved teaching. Thus, CU Boulder seeks to address the calls to professionalize teaching and create a climate of continuous improvement. The framework defines teaching as a scholarly activity—like research—and assesses the core components of such scholarship (Bernstein, 2008). Current efforts draw from decades of research in teaching evaluation to create a common framework (Glassick *et al.*, 1997) by defining categories of evaluation as follows: 1) clear goals, 2) adequate preparation, 3) appropriate methods, 4) significant results, 5) effective presentation, and 6) reflective critique.

These framework categories are held constant across all departments; however, specific interpretation of the components of the framework and their relative weights are defined at a department level (Figure 3). Thus, departments specify in a clear way what is meant by “multiple measures” and “significant results” locally, but use common categories across the campus.

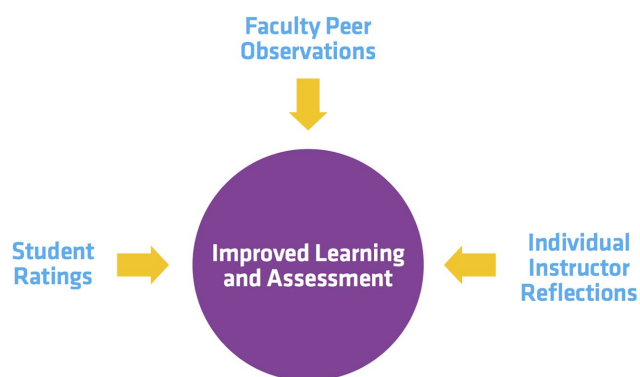


FIGURE 3. Three “voices” in a scholarly framework for assessing teaching.

This approach provides the university with a common framework while preserving disciplinary identity and specificity.

The implementation strategy has created two layers of work: one at the departmental level and one at the campus-wide level. Participation in the TQF is purposefully voluntary, asking departments to work to develop the framework rather than address a top-down mandate. The departmental level seeks to increase engagement and exploration of new ways to assess teaching by empowering individual departments to identify how they might enact more scholarly measures of teaching. Nine CU Boulder departments participated in the TQF in the 2016–2017 academic year with a postdoctoral-level facilitator. These cross-departmental discussions have led to departmental-based work in Fall 2017. An initial four departments have committed two to five lead faculty to identify what the measures of scholarly teaching are that address the framework in their disciplines, while the facilitator manages the biweekly meetings and shares information across departments. Additional departments are expected to phase in participation in Spring and Fall 2018.

CU Boulder has plans for two levels of campus discussions: the first among the pilot departments; and a subsequent one that will include broader representation from other departments, deans, and other institutional stakeholders. Once the departmental metrics and common campus framework and review system are coordinated, these tools will be deployed in the annual merit review and/or promotion and tenure review of departments across campus.

CONCLUSION

There is no question that strong examples of excellent teaching practice already exist throughout research universities. However, increasing visibility of and institutionalizing support for and reward of effective teaching is a challenge faced by many research universities. In most cases, relevant policies are already in place that emphasize the importance of teaching, but work remains to change the culture such that common practice aligns with these policies, especially at the departmental level.

Here, we have outlined some key elements associated with reward structures within research universities that can be leveraged to align practice and policy. To illustrate potential variations within the general framework, we highlighted three different approaches that are being piloted at specific research universities. At the department level, there needs to be an explicit conversation about the scholarly nature of teaching and a faculty member’s responsibilities regarding teaching as a scholar in a particular discipline. There also needs to be an explicit discussion of the collective nature of undergraduate teaching and its role within the broader responsibility of the research university. Finally, there needs to be recognition and adoption of empirical models for evaluating teaching that have been tested and validated. Within this broader context, the specific implementation at any given research university must be flexible and adaptable to local culture, structures, and goals.

ACKNOWLEDGMENTS

This research collaborative between the AAU and the Cottrell Scholars is supported by funds from the RCSA.

REFERENCES

- Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching*. San Francisco: Wiley.
- Anderson, W. A., Banerjee, U., Drennan, C. L., Elgin, S. C. R., Epstein, I. R., Handelsman, J., ... Warner, I. M. (2011). Changing the culture of science education at research intensive universities. *Science*, *331*(6014), 152–152. <https://doi.org/10.1126/science.1198280>
- Becker, E., Easlson, E., Potter, S., Guzman-Alvarez, A., Spear, J., Facciotti, M., ... Pagliarulo, C. (2015). *The Effects of Practice-Based Training on Graduate Teaching Assistants' Classroom Practices*. bioRxiv. <https://doi.org/10.1101/115295>
- Bernstein, D. (2008). Peer review and evaluation of the intellectual work of teaching. *Change: The Magazine of Higher Learning*, *40*(2), 48–51.
- Bernstein, D. J., & Huber, M. T. (2006). What is good teaching? Raising the bar through Scholarship Assessed. Invited presentation to the International Society for the Scholarship of Teaching and Learning (Washington, DC).
- Boring, A., Ottoboni, K., & Stark, P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*. <https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1>
- Bradforth, S. E., Miller, E. R., Dichtel, W. R., Leibovich, A. K., Feig, A. L., Martin, J. D., ... Smith, T. L. (2015). University learning: Improve undergraduate science education. *Nature*, *532*(7560), 282–284. <https://doi.org/10.1038/523282a>
- Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, *41*, 71–88. <https://doi.org/10.1016/j.econedurev.2014.04.002>
- Brown, P. C., Roediger, H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*. Cambridge, MA: Harvard University Press.
- Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *Journal of Higher Education*, *71*(1), 17–33.
- Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education*, *31*(1), 16–30.
- Dancy, M., & Henderson, C. (2010). Pedagogical practices and instructional change of physics faculty. *American Journal of Physics*, *78*(10), 1056–1063.
- Dancy, M., Henderson, C., & Smith, J. (2014). Understanding educational transformation: Findings from a survey of past participants of the Physics and Astronomy New Faculty Workshop. In *Proceedings of the 2013 Physics Education Research Conference* (pp. 113–116).
- Doyle, T. (2008). *Helping students learn in a learner-centered environment: A guide to facilitating learning in higher education*. Sterling, VA: Stylus.
- Eddy, S., & Hogan, K. (2014). Getting under the hood: How and for whom does increasing course structure work? *CBE—Life Sciences Education*, *13*(3), 453–468. doi: 10.1187/cbe.14-03-0050
- Fairweather, J. (2002). The ultimate faculty evaluation: Promotion and tenure decisions. *New Directions for Institutional Research*, *114*, 97–108.
- Fairweather, J. (2009). Work allocation and rewards in shaping academic work. In Enders, J., & deWeert, E. (Eds.), *The changing face of academic life: Analytical and comparative perspectives* (pp. 171–192). New York: Palgrave Macmillan.
- Fairweather, J., & Beach, A. (2002). Variation in faculty work within research universities: Implications for state and institutional policy. *Review of Higher Education*, *26*(1), 97–115.
- Finkelstein, N., Reinholz, D. L., Corbo, J. C., & Bernstein, D. J. (2015). *Towards a teaching framework for assessing and promoting teaching quality at CU-Boulder* (Report from the STEM Institutional Transformation Action Research [SITAR] Project). Boulder, CO: Center for STEM Learning. Retrieved January 1, 2016, from www.colorado.edu/csl/au/resources/TQF_WhitePaper_2016-1-17.pdf
- Freeman, S., Eddy, S., McDonough, M., Smith, M., Okoroafor, N., Jordt, H., & Wenderoth, M. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences USA*, *111*(23), 8410–8415. doi: 10.1073/pnas.1319030111
- Glassick, C. E., Huber, M. T., & Maeroff, G. I. (1997). *Scholarship assessed: Evaluation of the professoriate*. San Francisco: Jossey-Bass.
- Gormally, C., Evans, M., & Brickman, P. (2014). Feedback about teaching in higher ed: Neglected opportunities to promote change. *CBE—Life Sciences Education*, *13*(2), 187–199.
- Haak, D., Hille Ris Lambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science*, *332*(6034), 1213–1216. doi: 10.1126/science.1204820
- Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., ... Wood, W. (2004). Policy forum: Scientific teaching. *Science*, *304*(5670), 521–522.
- Henderson, C., & Dancy, M. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics—Physics Education Research*, *3*(2), 020102.
- Huber, M. T. (2002). Faculty evaluation and the development of academic careers. *New Directions for Institutional Research*, *114*, 73–84. doi: 10.1002/ir.48
- Hutchings, P. (Ed.) (1995). *From idea to prototype: The peer review of teaching*. Sterling, VA: Stylus.
- Hutchings, P. (Ed.) (1996). *Making teaching community property: A menu for peer collaboration and peer review*. Sterling, VA: Stylus.
- Kober, N. (2015). *Reaching students: What research says about effective instruction in undergraduate science and engineering*. Washington, DC: National Academies Press.
- Lorenzo, M., Crouch, C. H., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. *American Journal of Physics*, *74*(2), 118–122. doi: 10.1119/1.2162549
- Lowen, R. S. (1997). *Creating the Cold War university: The transformation of Stanford*. Berkeley: University of California Press.
- Lyde, A. R., Grieshaber, D. C., & Byrns, G. (2016). Faculty teaching performance: Perceptions of a multi-source method for evaluation (MME). *Journal of the Scholarship of Teaching and Learning*, *16*(3), 82–94.
- MacNeil, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, *40*, 291. <https://doi.org/10.1007/s10755-014-9313-4>
- Malcolm, S. & Feder, M. (Eds.) (2016). *Barriers and opportunities for 2-year and 4-year STEM degrees: Systemic change to support students' diverse pathways*. Washington, DC: National Academies Press. doi: 10.17226/21739
- National Research Council. (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). Washington, DC: National Academies Press. doi: 10.17226/9853
- Singer, S. (2015). Keynote: Implementing evidence-based undergraduate STEM teaching practice. In *Searching for better approaches: Effective evaluation of teaching and learning in STEM* (pp. 1–5). Tucson, AZ: RCSA. Retrieved July 15, 2015, from <https://rescorp.org/gdresources/publications/effectivebook.pdf>
- Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (Eds.) (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, DC: National Academies Press.
- Trenshaw, K. F., Targan, D. M., & Valles, J. M. (2016). Closing the achievement gap in STEM: A two-year reform effort at Brown University. In *Proceedings from ASEE NE '16: The American Society for Engineering Education Northeast Section 2016 Conference*. Retrieved August 24, 2017, from <https://egr.uri.edu/wp-uploads/asee2016/73-1064-1-DR.pdf>
- Wieman, C. (2015). A better way to evaluate undergraduate teaching. *Change: The Magazine of Higher Learning*, *47*(1), 6–15. doi: 10.1080/00091383.2015.996077