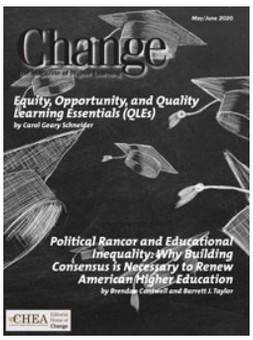

Routledge
Taylor & Francis Group

Change: The Magazine of Higher Learning

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/vchn20

Establishing a Better Approach for Evaluating Teaching: The TEval Project

Gabriela C. Weaver , Ann E. Austin , Andrea Follmer Greenhoot & Noah D. Finkelstein

To cite this article: Gabriela C. Weaver , Ann E. Austin , Andrea Follmer Greenhoot & Noah D. Finkelstein (2020) Establishing a Better Approach for Evaluating Teaching: The TEval Project, Change: The Magazine of Higher Learning, 52:3, 25-31, DOI: 10.1080/00091383.2020.1745575

To link to this article: https://doi.org/10.1080/00091383.2020.1745575

Published online: 21 Jul 2020.

Establishing a Better Approach for Evaluating Teaching: The TEval Project

By Gabriela C. Weaver, Ann E. Austin, Andrea Follmer Greenhoot, and Noah D. Finkelstein

Gabriela C. Weaver is Professor of Chemistry at the University of Massachusetts Amherst. She also serves as Special Assistant to the Provost for Educational Initiatives. Her research focuses on developing and assessing approaches to give students more authentic learning experiences and how to transform teaching at the faculty, departmental, and campus levels to achieve greater student success.

Ann E. Austin is University Distinguished Professor of Higher, Adult, and Lifelong Education, Associate Dean for Research in the College of Education, and Assistant Provost for Faculty and Academic Staff Development at Michigan State University. Her research concerns organizational change in higher education; science, technology, engineering, and math education; and academic careers.

Andrea Follmer Greenhoot is Professor of Psychology, Director of the Center for Teaching Excellence, and Gautt Teaching Scholar at the University of Kansas. Her research focuses on memory and cognitive development and on strategies for advancing transformed teaching and learning, informed by cognitive and developmental science in higher education.

Noah D. Finkelstein is Professor of Physics and Co-Director of the Center for STEM Learning at the University of Colorado Boulder. He conducts research in physics education,specifically studying the conditions that support students' interests and abilities in physics. In parallel, he conducts research on how educational transformations get taken up, spread, and are sustained.

In Short

- Student surveys of courses are at best unreliable or at worst discriminatory methods to evaluate the quality of teaching.
- Teaching activities include many aspects that are not visible in the classroom that require a more holistic approach to evaluation.
- The Transforming the Evaluation of Teaching project is a multi-institution collaboration to refine and implement a teaching evaluation framework that addresses the shortcomings of typical approaches.
- Our cross-case analysis examines the ways that institutional context affects the cultural change process of adoption of this new approach, one of numerous efforts across the country to determine best practices around teaching evaluation.

Effective teaching practice focused on encouraging deep engagement helps all students learn and succeed. Considerable work by researchers and change agents, encouraged by national organizations such as the National Science Foundation (NSF) and the National Research Council, has promoted greater use of evidence-based educational practices, such as peer learning, the use of authentic assignments, and problem-based learning known to improve learning and reduce disparities in outcomes across learners.

However, although we know much about what effective teaching looks like and the power of its impact, universities and colleges have found that changing teaching practice is challenging. Faculty members are informed by the more traditional ways in which they themselves learned, and promotion and tenure processes typically emphasize research accomplishments over teaching achievements.

With these pressures in mind, one way to change faculty teaching behaviors is through explicit recognition and reward of those who use effective practices. However, doing so requires evaluation processes that better align to those practices. Evaluation and reward systems send messages about what is valued in the organizational culture. When these systems are aligned with research on teaching and learning, it opens avenues for discussion among colleagues about ideas and possibilities for improving teaching practice.

The common approach to evaluating teaching in U.S. higher education, including in the 2-year college sector, involves a questionnaire answered by students of a course at the end of each term. The exact questions on the surveys, the delivery and incentive processes, and the response rates vary somewhat from institution to institution, but the basic approach is the same.

The ease of administering and compiling the results from such surveys across disciplines contributes to their appeal, but criticisms and concerns about this approach are mounting. (Basow & Martin, 2012; MacNell, Driscoll, & Hunt, 2015; Reinsch, Goltz, & Hietapelto, 2020; Spooren, Brockx, & Mortelmans, 2013).

Faculty members themselves often express dissatisfaction when student ratings are the sole vehicle for assessing teaching quality. They suggest that the full spectrum of their teaching activities is not adequately represented through these ratings. They recognize that their work and the quality of the learning experiences they cultivate relate to many things—their approaches to course design, their choice of learning resources, the assignments they design, the ways they interact with students, their processes of reflection on their teaching and their students' learning, and their efforts to make adjustments in their teaching.

However, typical approaches to evaluating teaching through student evaluations are not designed to capture this array of work nor to foster reflection and conversation that translates evaluation into

improvement. More holistic approaches are needed that include attention to the various aspects of teaching and take into account different approaches to teaching within the range of disciplinary contexts.

A recent study, which used a computer simulation to examine the ability for student evaluation scores to provide information about the quality of teaching, found that "even under ideal circumstances … SETs [student evaluations of teaching] still yield an unacceptably high error rate" (Esarey & Valdes, 2020). The authors conclude that "a combination of independent evaluators, interviews with students, teaching observations by experts, peer review of instructional materials and SET scores can give a much more accurate picture of a faculty member's teaching proficiency when SET scores alone would be misleading" (p. 3).

The interest in discussing teaching evaluation and creating more effective approaches is evident across the country, as well as internationally. Conversations and initiatives are being explored at many higher education institutions. At the national level, the NSF is funding a project which we lead, on Transforming the Evaluation of Teaching (or TEval, teval.net, described in detail below), involving multiple universities engaged in institutional change to advance science, technology, engineering, and math (STEM) undergraduate education through efforts to create more effective approaches to teaching evaluation. The Association of American Universities (AAU) is focusing on this issue as part of its STEM Initiative, and the National Academies of Sciences, Engineering, and Medicine hosts a Roundtable on Systemic Change in Undergraduate STEM Education that, as part of its work, is focusing on teaching evaluation.

If the culture within higher education institutions is to value teaching more highly and if teaching practice is to reflect more fully the lessons from the growing body of research on teaching and learning, then teaching evaluation must be improved. In this article, we describe an approach to transforming the institutional culture around teaching through the development, adoption, and sustainable use of new approaches to teaching evaluation.

We define holistic evaluation, describe ways that departments are engaged in the process of changing their approaches to teaching evaluation, and highlight lessons about how to engage in this kind of organizational change work. As higher education institutions experiment and share ideas about approaches to changing the evaluation and reward system around teaching, they are promoting significant organizational change in support of better teaching that serves the learning needs of the full array of student populations.

Structure of the TEval Project

The TEval Project seeks to not only develop and implement a better approach to teaching evaluation, but also to study the process of adoption and change within the higher education institutional culture. To that end, our study is a collaboration among four public, land-grant, high research productivity universities: Michigan State University (MSU), the University of Colorado Boulder (CU), the University of Kansas (KU), and the University of Massachusetts, Amherst (UMass). Three of these institutions are the implementation sites: KU, CU, and UMass. These three institutions are refining and implementing a holistic approach to teaching evaluation based on common principles but with context-sensitive variations in approach. The fourth team, MSU, is carrying out a cross-case analysis of the three campus sites with the goal of understanding the change process, especially how that change process cuts across institutional settings or may be specific to each.

The end-of-term survey approach to teaching evaluation provides only one lens on the practice of teaching—the perception of students who are enrolled in a course—and it is based on only one form of evidence—a survey. A less frequently utilized but somewhat common additional form of teaching evaluation involves peer observation. This often takes the form of another faculty member from the same department being an observer for one, or occasionally a few, class sessions of a fellow faculty member.

In some instances, peer evaluation has begun to be based on observation instruments that provide guidance for observers' note taking and reporting. Depending on the validity of the observation instrument used and the training of the observer, if any, this form of peer evaluation can result in a narrow description and benchmark-free opinion of the observed teaching. This form of peer evaluation is also based on one lens—an external observer—and one dimension of teaching—in-class activities.

The framework that the TEval partner institutions have decided to implement assumes that the practice of teaching involves many dimensions, a number of which are not observable in the classroom. The framework's components are based on two decades of scholarship on teaching and its evaluation (Bernstein & Huber, 2006; Glassick, Huber, & Maeroff, 1997; Hutchings, 1995, 1996; Lyde, Grieshaber, & Byrns, 2016) and related work on the peer review of teaching (Bernstein, 2008).

From this literature, the work on each campus centers on the development and use of a teaching evaluation rubric that we believe is appropriately holistic, providing a richer, more complete view of teaching practice and the evidence that speaks to it. Although the emergent rubrics and processes at each campus differ in details, the foundational framework of each is based on these principles:

- ***Evaluation includes multiple dimensions of teaching***—categories of activities that capture the teaching endeavor in its totality, including aspects that take place outside of the classroom.
- ***Evaluation includes multiple lenses***—multiple sources and types of data, including faculty self-report (e.g., course materials, evidence of student learning, and reflections on it), peer input (e.g., class visits, review of course materials, discussions with the instructor), and student voices (e.g., course evaluations, alumni letters).
- ***Evaluation involves triangulation***—no measure should be used in isolation, and measures should support each other.
- ***Both formative and summative*** uses of the data must be possible to maximize the impact on teaching effectiveness.
- There must be a ***balance between uniformity across departments and customization*** for different disciplines in order to maximize usefulness at the institutional level.

Work that originated with the KU team before the multi-institutional collaborative came together had already identified the seven dimensions that have become integrated into the TEval framework: goals, content, and their alignment; teaching practices; achievement of learning outcome goals; reflection and iterative growth; mentoring and advising; classroom climate and student perceptions; and involvement in service, scholarship, or community related to teaching (Follmer Greenhoot, Ward, & Bernstein, 2017).

The resulting rubric for evaluation articulates what effective and inclusive practices look like within each dimension of teaching. In addition, each dimension is intended to be evaluated by using more than one lens or source of information (student, instructor, third party) and multiple types of evidence (quantitative surveys, focus group discussions, observations, document analysis, etc.), as shown schematically in Figure 1. Each dimension can be evaluated along a scale to provide quantitative information or with written input to provide qualitative feedback.

The multi-institutional collaborative explored scalable change processes in which the department as a whole would serve as the unit of change.

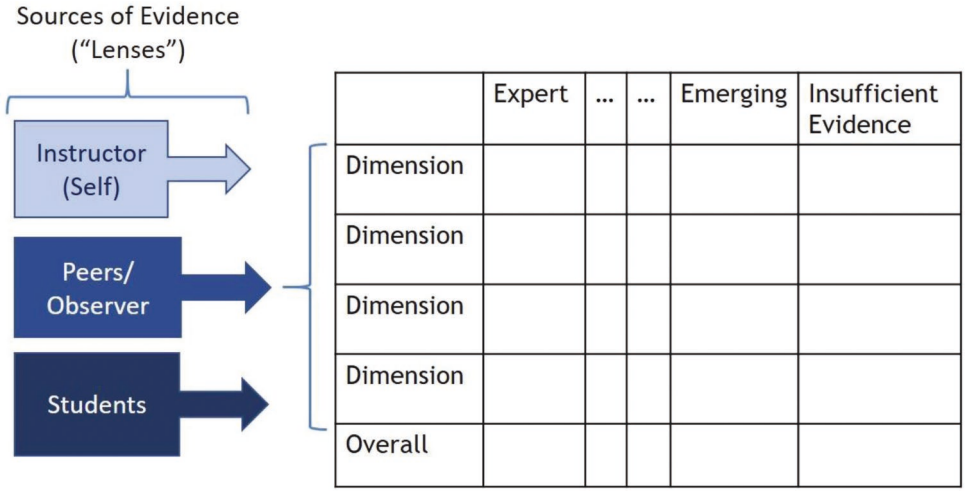

Figure 1. Generic Version of a Holistic Teaching Evaluation Rubric

Although projects that seek to change pedagogies or curricula often focus on or originate with a single faculty member (National Research Council, 2012), working with the entire department as the unit of change supports institutionalization and integration with policies and practices at the broader college, school, or institutional levels.

The processes adopted in TEval are intended to support departments as learning communities that form a shared vision of effective teaching within the department, which is the primary unit for faculty evaluation and academic organization. Departments opting to participate are involved in both the development and customization of the rubric and in piloting it. They also ultimately choose what the form of implementation will look like—for example, whether focusing on a formative process, an evaluative one, a process tailored to pre-tenure faculty, or one that will involve all faculty but only every second or third year.

The multi-institutional collaborative also brings with it the benefit of knowledge sharing. Once yearly, the project brings together faculty from departments at each of the implementation institutions along with project leaders from all four campuses at Knowledge Exchange meetings in which they are able to problem solve with and learn from each other. In some cases, the faculty participants are able to speak with faculty of the same, or similar, disciplines at another institution. But the entire gathering benefits from seeing how the same overall process plays out in slightly different settings (disciplines and/or institutions), with that diversity of experiences prompting creative solutions to implementation challenges.

Department-Level Activities

At each implementing institution, the process has some slight variations. For each of the first 3 years of the project, the project team at the institution issued a call for participation to departments (see Table 1).

Table 1. Participating Departments at Each Implementation Site

Year 1	Year 2	Year 3
University of Colorado Boulder		
Integrative Physiology	*Germanic and Slavic Languages and Literatures	College of Engineering and Applied Science (5–11 additional departments/units)
Mathematics	*History	
Mechanical Engineering	Residential Academic Programs	
Molecular, Cellular, and Developmental Biology		
*School of Business		
University of Kansas		
Chemical and Petroleum Engineering	*African and African American Studies	Civil, Environmental, and Architectural Engineering
*Philosophy	*French, Italian, and Francophone Studies	*Pharmacy Practice
*Public Affairs and Administration	Linguistics	*School of Social Welfare
Sociology	Physics and Astronomy	
Undergraduate Biology		
University of Massachusetts, Amherst		
Environmental Conservation	Anthropology	*Art
Information and Computer Sciences	Biochemistry and Molecular Biology	Nutrition
*Linguistics	*Judaic and Near Eastern Studies	
*Music and Dance		
Physics		

*Non-STEM departments/schools have been supported with institutional funds to match NSF funds supporting STEM departments.

The call differed across the implementing institutions but included common elements: that each department will participate for at least two years, it will attend meetings to work on the rubric design and implementation plans, and in turn, it will receive both financial and human support.

The financial support is a small discretionary fund ($3–$5K) given to each department to support activities related to the planning and implementation activities. The human support may involve a team of pedagogical experts working closely with a departmental action team, as is done at CU Boulder (Reinholz, Corbo, Dancy, & Finkelstein, 2017) or cross-departmental meetings interspersed with department-specific meetings with TEval leaders on an as-needed basis.

In each participating department, the work of transforming teaching evaluation is intended to be a department-wide effort. However, in the initial stages, the departmental working groups involved three to five faculty members who are very interested in teaching issues. This process included some faculty members who are either not tenure-stream and/or who teach many of the lower-level courses. It was also helpful to have one or two who are senior-level faculty respected among the research-focused faculty but with recognized interest in excellence in teaching.

In addition, the head or chair is either part of the working group—as is a requirement of participation at UMass—or is a supporter of the effort but involved at the level of department-wide action. It has been particularly effective when some faculty in the department (both tenure-stream and not) have had some level of engagement with teaching development through their campus, since these faculty can help to link the pedagogical discussions to the culture of their department.

Although the dimensions of teaching have been defined by the framework itself, departments have needed to build consensus about how these are each expressed and enacted in their discipline—what types of learning goals, pedagogies, mentoring, and external activities make sense and are expected. Departments also explore variations in the weighting that each teaching dimension has in the evaluation process, as well as the benchmarks that are expected for instructors at different career levels.

The discussions leading to this consensus are an important part of changing the culture around teaching and teaching evaluation. They help to generate shared understanding about the role of teaching in the mission of the department and the functions of the curriculum as a whole. The working group begins this process, but it must involve the entire faculty of any department in order to be implemented and sustained. Therefore, the TEval project leaders support the working groups in engaging their faculty colleagues department-wide in these consensus-building discussions.

As expected, different departments at each of the institutions have evolved different approaches to adapting and adopting the holistic framework for teaching evaluation. Some have developed mentoring or evaluation teams (dyads or triads) that work with a faculty member being evaluated, whereas others have a single departmental committee that will do the evaluation work for all who are being evaluated. In some departments, using this form of evaluation begins with a voluntary group of faculty whereas other departments begin with faculty of a particular rank. Some departments are choosing to implement this approach for a purely formative purpose whereas others are transitioning to an evaluative purpose after a short piloting phase. The framework for TEval enables this level of customization for departmental needs and disciplinary cultures, while simultaneously providing a uniform foundation that can be utilized across a campus.

Middle-Out Change: Engaging the Administration

The TEval project is structured for middle-out institutional transformation. Working at the department level allows individual faculty members to make changes to their teaching practices and beliefs and also supports campus-level changes by leveraging the experiences of multiple departments that are joined in the effort (Reinholtz et al., 2015). As described previously, one long-range goal for improving the evaluation of teaching is to facilitate a more consistent inclusion of teaching into university reward systems at a higher level than is currently the norm.

For this to happen, members of the administration—at least as high as the provost and including deans—as well as faculty governance bodies (Senate and unions, if they apply) must be prepared to accept and support the change originating with the departments. The TEval project leaders at each campus

have engaged with the deans, provost, and other administrators in a variety of ways from the beginning of the project, describing the work of the departments and seeking to pave the way for scaling up once the work has reached a mature level. In this way, working first with departments influences both the individual faculty (the "grass roots") and the administration (the campus level) for middle-out change.

Observations and Emerging Lessons

As we interact within and across these universities, we are learning that faculty members and institutional leaders are motivated to engage in work that enables them to grapple with defining "good teaching" within their departmental and institutional contexts. They also value making the various dimensions of their teaching efforts more visible, highlighting different forms of teaching excellence, and elevating the value of teaching as part of their institutional missions.

At the same time, we see the challenges and barriers that arise in this work. While those faculty and administrators involved in the work often report that it is stimulating and rewarding, they also recognize that it requires an allocation of time in busy schedules. Gaining wide commitment and buy-in from faculty members across a department requires attention and, at some institutions, attracting the investment and commitment of busy senior-level administrators may also be a challenge. We hope that departments will need to invest less time when they build on and adapt the tools and processes developed by departments that have done the pilot work.

Another issue is knowing where to begin work on teaching evaluation issues. Some departments start with informal discussions among faculty about concerns with student ratings; others begin by considering the elements of faculty assignments. The work may also need to deal with uncertainty about whether and how broader forms of evidence of teaching quality and outcomes will be integrated into institution-level tenure and promotion processes.

While acknowledging challenges associated with this work, we also have learned about strategies that project leaders at the institutional and departmental levels are using to make progress. Since changing the way in which teaching is evaluated and valued is an organizational goal, taking a systemic approach is wise, using common strategies like these:

***Examining the landscape and its potential relevance to changing teaching evaluation*:** Each institution and department has a particular culture and history. Taking time to consider those histories and contextual features can be useful in making decisions about how to change approaches to teaching evaluation. The Center for Teaching Excellence (CTE) at KU is well respected across campus; thus, situating their teaching evaluation project within the CTE invites interest and confidence.

***Establishing and articulating a vision*:** Some institutions and departments have created organizational narratives about new approaches to teaching evaluation that connect to other institutional narratives. For example, UMass leaders articulate how efforts to improve approaches to teaching evaluation relate to institutional priorities to identify learning outcomes across the curriculum. Similarly, KU emphasizes its prior work on making teaching visible and has linked efforts to rethink teaching evaluation to that earlier work.

Building capacity and engagement: In order to foster change in teaching evaluation, institutional and departmental leaders need to find ways to attract and engage a broad range of faculty members. Our case institutions utilize a range of effective strategies: showing faculty how efforts to change teaching evaluation relate to their own personal concerns around fair tenure and promotion processes, tapping into faculty concern about the limitations and biases that can appear in traditional student evaluations of teaching, and linking to faculty members' natural interest in teaching and collegiality. One department chairperson highlighted the "crisis of the humanities" and the associated patterns of dropping enrollments; the chair continued by reminding faculty colleagues that excellent teaching helps attract students to the department, and attending to the evaluation of teaching may increase the department's reputation for caring about good teaching.

***Gaining the support of leaders at various levels of the institution*:** Efforts to change how teaching is valued and evaluated involves an organizational change process. In such work, the support and encouragement of change champions at the department, college, and institutional levels can have an important impact. The middle-out approach allows organizational change processes to leverage both "top-down" and "bottom-up" leadership simultaneously.

***Embedding the work of discussing and changing teaching evaluation into "regular" departmental work*:** In some departments at our case universities, leaders have integrated discussion and work associated with teaching evaluation into regularly scheduled meetings. In this way, many colleagues become involved in the work, and the time needed has already been allocated in busy schedules. Symbolically, faculty can see that time spent on teaching issues is time valued in the life of the department.

***Identifying allies*:** Organizational change efforts benefit from wide support. One way to achieve such support is to identify potential allies likely to be interested in changing traditional approaches to teaching evaluation. For example, senior institutional leaders working on increasing student academic success may recognize that new approaches to teaching evaluation could aid their work. Other allies may be departmental colleagues who value equity, diversity, and inclusion and who see new approaches to documenting and assessing the elements of teaching practice as a path toward increased recognition of the diverse ways in which individual faculty members contribute to the collective good.

In addition to the cross-case analysis research carried out by the MSU team, the TEval project is undergoing external evaluation by a team led by Mark Graham at Yale University. Theirs is an innovative approach to project evaluation using collaboratively generated visual process maps of planned and actual work over time. Using this approach, combined with the case study research, we are observing a range of strategies across the departments engaged in discussion and experimentation around teaching evaluation. We expect to learn and share more about strategic options that departments might consider as they continue with this work.

Looking Forward

The work of the TEval Project is one element in a growing national movement, especially in the STEM fields, to improve teaching evaluation and strengthen undergraduate learning. The TEval project provides a network within which the three participating universities each engage with teaching evaluation reform in ways consistent with their institutional histories and contexts—while also working together to share strategies, compare progress, and collectively reflect and learn.

In addition to the work underway through the TEval project, other national initiatives are also gaining traction. The National Academies of Sciences, Medicine, and Mathematics has appointed a Roundtable on Systemic Reform in STEM Undergraduate Education. The Roundtable's mission is to encourage systemic efforts to improve teaching and learning in STEM undergraduate education by linking and connecting projects and initiatives with similar goals and identifying issues and gaps still to be addressed in support of deeper STEM undergraduate learning.

Supporting the improvement of teaching evaluation is one of the priority areas of the Roundtable. The Roundtable, along with the TEval team and the Association of American Universities' STEM Initiative, hosted a 2-day workshop in September 2019 on "Recognizing and Evaluating Science Teaching in Higher Education." The conversations of the approximately 40 participants focused on the history of teaching evaluation, efforts currently underway to improve it in a range of institutional types, how reform of teaching evaluation is part of systemic change efforts to strengthen teaching, and next steps to advance these reform goals. A proceedings brief is available (National Academies of Sciences, Engineering, and Medicine, 2020). The Academies' Roundtable, working with AAU, Accelerating Systemic Change Network, and the TEval team, also is developing plans to host a national convening in the coming year on the topic of improving teaching evaluation in STEM undergraduate education.

Examining and rethinking teaching evaluation processes is a strategic lever to achieve more effective teaching and deeper student learning. The TEval approach is adaptable to other institutions because it approaches faculty and their departments as the key locus for changing teaching and evaluation practices. This approach enables the faculty in each department to decide the specific challenges they face and the paths forward that they find more relevant and potentially effective for their situation.

As faculty grapple with what constitutes good teaching, what is the full array of work activities that encompass teaching practice, and what evidence and artifacts represent their work as teachers, they are engaged in the work of placing teaching at the heart of work in higher education. These efforts contribute to organizational and cultural change that ultimately advances a primary mission in American higher education—the learning of our students. C

This material is based upon work supported, in part, by the National Science Foundation, Award numbers: DRL 1725946, 1726087, 1725959, and 1725956. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. TEval is an initiative of the Bay View Alliance, an international network of research universities exploring strategies to support and sustain the widespread adoption of instructional methods that lead to better student learning. https://bayviewalliance.org

References

- Basow, S. A., & Martin, L. J. (2012). Bias in student evaluations. In M. E. Kite, *Effective evaluation of teaching: A guide for faculty and administrators* (pp. 40–49). Washington, DC: Society for the Teaching of Psychology. Retrieved from http://teachpsych.org/ebooks/evals2012/index.php

- Bernstein, D. (2008). Peer Review and Evaluation of the Intellectual Work of Teaching. *Change: The Magazine of Higher Learning, 40*(2), 48–51. Retrieved from http://doi.org/10.3200/CHNG.40.2.48-51

- Bernstein, D., & Huber, M. T. (2006). What is good teaching? Raising the bar through Scholarship Assessed. In *International Society for the Scholarship of Teaching and Learning*. Washington, DC.

- Esarey, J., & Valdes, N. (2020). Unbiased, reliable and valid evaluations can still be unfair. *Assessment and Evaluation in Higher Education.* https://doi.org/10.1080/02602938.2020.1724875

- Follmer Greenhoot, A., Ward, D., & Bernstein, D. (2017). *Benchmarks for teaching effectiveness.* Retrieved from https://cte.ku.edu/rubric-department-evaluation-faculty-teaching

- Glassick, C. E., Huber, M. T., & Maeroff, G. I. (1997). *Scholarship assessed: Evaluation of the professoriate. Special report*. San Francisco, CA: Jossey-Bass.

- 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.

- 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. https://doi.org/10.14434/josotl.v16i3.18145

- MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, *40*(4), 291–303. https://doi.org/10.1007/s10755-014-9313-4

- National Academies of Sciences, Engineering, and Medicine. (2020). *Recognizing and evaluating science teaching in higher education: Proceedings of a Workshop—In Brief*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25685

- National Research Council. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, DC: National Academies Press. https://doi.org/10.17226/13362

- Reinholz, D. L., Corbo, J. C., Dancy, M. H., Deetz, S., and Finkelstein, N. (2015). Towards a Model of Systemic Change in University STEM Education. In G. C. Weaver, W. D. Burgess, A. L. Childress, and L. Slakey (Eds.), Transforming Institutions: 21st Century Undergraduate STEM Education. West Lafayette, IN: Purdue University Press.

- Reinholz, D. L., Corbo, J. C., Dancy, M., & Finkelstein, N. (2017). Departmental Action Teams: Supporting faculty learning through departmental change. *Learning Communities Journal*, *9*, 5–32.

- Reinsch, R. W., Goltz, S. M., & Hietapelto, A. B. (2020). Student evaluations and the problem of implicit bias. *Journal of College and University Law*, *45*(1), 114–140.

- Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, *83*(4), 598–642. https://doi.org/10.3102/0034654313496870