



Like 118

Share

Tweet

Share

154

## How Do We Evaluate Teaching?

*Findings from a survey of faculty members.*

By Craig Vasey and Linda Carroll

In fall 2014, the AAUP's Committee on Teaching, Research, and Publication conducted a survey to gather information about how colleges and universities evaluate teaching and use the results. The committee hoped that this survey would help faculty members improve evaluation practices at their institutions and enable them to defend themselves and their colleagues more effectively against the misuse of student evaluations. A few respondents explicitly expressed the hope that the AAUP would provide a list of best practices or recommended guidelines.

The survey was sent by e-mail to approximately 140,000 faculty members, and 9,314 responses were received. Although these responses were informative and useful, the survey did have imperfections. While it was intended to reach all faculty members regardless of status, the response from non-tenure-track faculty members was lower than the committee had hoped: 75 percent of the respondents were tenured or tenure-track faculty members.

At the same time, the volume of the responses received—within about eight weeks, nine thousand responses had come in, including five thousand written comments—indicates that the evaluation of teaching is a matter of concern for many faculty members. The fact that the majority of the respondents were in relatively advantaged tenured and tenure-track positions and still felt their lives were being affected by evaluation practices likely means that the effects are even worse for those off the tenure track. It seems highly unlikely that the comparatively low response rate from their more vulnerable colleagues is an indication that the issue doesn't matter to non-tenure-track faculty members. We need to find ways to hear more from this new majority.

### Overview of Findings

The issues at stake in evaluation practices may seem insignificant, but they can affect our lives in big ways. Through the survey we sought to collect information about how student evaluations of teaching performance are handled and used in salary, promotion, and tenure decisions; how various institutions balance the responsibilities of research and teaching; and what kind of support for teaching is available at different institutions.

Of the responses received, 54 percent came from tenured professors, 18 percent from full-time non-tenure-track faculty members, 15 percent from tenure-track faculty members, 11 percent from parttime non-tenure-track faculty members, and just over 1 percent from teaching or research assistants. Two hundred respondents did not identify their appointment type. Almost half of the responses came from faculty members at four-year teaching-intensive institutions (48 percent), followed by four-year research-intensive institutions at 35 percent. The rest were divided almost equally between two-year colleges and professional schools, including colleges within larger institutions.

The survey began with questions about mechanisms for evaluation of teaching and whether those mechanisms were recommended, required, required frequently or occasionally, or not recommended or required at all. The percentage of faculty members reporting frequent required use of paper student evaluations and the percentage reporting frequent required use of online student evaluations were almost identical at 51 and 52 percent, respectively. Very few respondents said that the student evaluations were recommended but not required (4 percent for paper and 9 percent for online). The same was true for the occasional use of required student evaluations (5 percent for paper and 7 percent for online). The required use of quantitative evaluations beat out required qualitative evaluations, but not by much (55 to 44 percent).

By an overwhelming margin, the responses regarding the shift from paper evaluations done in class to online evaluations done outside the classroom told the same story: the return rate has dropped from 80 percent or higher on paper to 20 to 40 percent online. With such a rate of return, claims of "validity" are rendered dubious. Faculty members reported that comments from the students are on the extremes: those who are very happy with their experience or their grade, and those who are very unhappy.

Some faculty members expressed frustration at having little to no input in determining what the evaluations ask, pointing out that it is inappropriate to treat all teaching in every field or all students as if they were the same. A common instrument takes no account of the differences between a lecture-based class delivered to more than fifty students and a seminar of fifteen.

Numerous reports indicated that the abusive and bullying tone often seen in anonymous online comments is beginning to appear in student evaluations. Some women faculty members and faculty members of color report receiving negative comments on appearance and

qualifications; it seems that anonymity may encourage such inappropriate and sometimes overtly discriminatory comments.

Most evaluations appear to be done in the last weeks of the semester; some schools allow them to be submitted even after students have received their grades. Commenters pointed out that factors such as stress and worry about grades increase for students toward the end of the semester, influencing their responses, and that allowing students to know their grades before evaluation occurs compromises the results by undermining objectivity.

About 25 percent of respondents said that their evaluations were frequently published—that is, made available to people other than the instructor and his or her department chair or dean. A majority (67 percent) said that their institutions did not require publication. About half said that, aside from student evaluations, they were evaluated frequently or occasionally by administrators, and about two-thirds said they were evaluated by peers.

Teaching portfolios, mentoring by faculty colleagues, or engagement with centers for teaching, while often recommended, were required only rarely. Although most respondents said that their institutions had centers for teaching (75 percent), few praised them for promoting better pedagogy. More often, faculty members associated them with efforts to promote technology or cater to students.

A small majority (55 percent) of respondents said that their institutions did not involve faculty members in decisions about the design of the evaluation instrument and its distribution. The gap grew when it came to the faculty's input concerning the use of student evaluations in promotion and tenure decisions and decisions on merit salary increases, with 62 percent saying that decisions about the use of evaluations did not lie with the faculty. The gap became even wider (65 percent) with regard to having input in decisions about publishing student evaluations.

The majority of the respondents (69 percent) saw a need for student evaluations of some sort, but respondents were more evenly split when it came to weighing their effectiveness: 50 percent said that student evaluations are not an effective means of determining good teaching, whereas 47 percent said that they are. Numerous commenters claimed that faculty members are evaluated and recommended for contract renewal or promotion on the basis of the grades they assign and that administrators pressure faculty members to pass students who deserve to fail. A majority (72 percent) strongly or somewhat strongly agreed that confidentiality—that is, the anonymity of the students completing the evaluations and some restrictions on who sees the resulting comments—was essential to the legitimate pedagogical purposes of student evaluations.

Most respondents recommended mentoring programs for junior faculty (86 percent); even more said that institutions should evaluate teaching as seriously as research and scholarship (90 percent). The majority said they believed that scholarship and research on teaching should be recognized as equal to disciplinary scholarship (84 percent).

Responses to a question about the use of outcomes assessment to improve teaching were almost evenly split among those who agreed that it is effective, those who said it is not effective, and those who had no opinion about its effectiveness. While the majority of respondents somewhat or strongly agreed (in equal measures) that student evaluations create upward pressure on grades (67 percent), 77 percent were opposed to the imposition of grade distribution quotas by the administration.

### The Contingent Faculty Perspective

Respondents cited significant differences between how administrators evaluate non-tenure-track faculty members and how they evaluate tenure-track and tenured faculty members. Most respondents also noted that the traditional means of providing oversight and support for teaching were limited to those on the tenure track. Non-tenure-track faculty members, including graduate students, receive significantly less support and often are excluded from participation in mentoring, teaching programs, instructional development, and peer evaluations. Given that non-tenure-track faculty members are responsible for teaching the majority of courses and that graduate students represent the next generation in higher education, this reported lack of mentoring and attention to quality seems surprising. It is challenging and disheartening to try to measure up to student and departmental expectations and to endure judgment of the quality of one's teaching under such circumstances.

For those on the tenure track, student evaluations may be important in the promotion and tenure processes; however, once tenured, some faculty members seem to accord evaluations little value. In contrast, for non-tenure-track faculty members the emphasis is on student evaluations rather than on innovative teaching, instructional development, or formal recognition of excellence in the classroom. Many commented that evaluations are used solely in the context of renewal or nonrenewal of contract. Others said that even highly favorable student evaluations do not usually change the status of non-tenure-track faculty members.

Rather than strengthen the quality of teaching, student evaluations and their use by institutions exacerbate the problems of a two-tier system, compromising the quality of education. Knowing that administrators rely on student evaluations in making renewal decisions, non-tenure-track faculty members face the challenge of balancing rigorous and interesting courses with the reality that many students prefer to maximize their grade point averages. Hastily completed evaluations, critical or complimentary, are soon forgotten by the students, who reduce to a few data points the months that faculty members have put into planning and teaching each course. All faculty members admit concerns about grade inflation and sustaining student interest, but the tenured faculty member's job is not in danger when a new course offering fails to attract a sufficient number of students.

In addition to the pressure to inflate grades in order to secure teaching assignments, contingent faculty members face pressure to raise course and program completion rates, which are tied to already much-reduced state funding. In the future, faculty members with no employment security likely will be under increasing pressure from both administrators and the expanding cadres of student-services staff hired to monitor students' progress.

## Recommendations from Respondents

Among the many open-ended comments submitted by respondents were a number of suggestions and recommendations. Here we summarize those that came up frequently.

The Association's *Statement on Teaching Evaluation* and the *Observations on the Association's Statement on Teaching Evaluation*, both of which are included in the eleventh edition of *AAUP Policy Documents and Reports*, provide the policy context for the committee's survey. The overwhelming majority of the recommendations that emerged from the responses comport with those made in the *Statement on Teaching Evaluation*, which calls for clear institutional policies, assessment by multiple parties, the use of instruments that are suited to the field of knowledge, and the use of evaluation for developmental purposes. AAUP policy documents also emphasize the primary role of the faculty in teaching evaluation and warn against the encroachment of "corporate forms of governance" and the growing reliance on numerically based evaluations.

Survey respondents echoed this emphasis on a strong role for the faculty, and particularly the faculty in the field, in determining the components and processes of teaching evaluation. Decisions made with little to no faculty involvement, such as the widespread move from paper to online forms, frequently have negative consequences such as those discussed above. Numerous commenters recommended that administrators give more weight to qualitative components (comments) rather than reducing evaluations to a number.

Respondents also recommended that the evaluation of teaching be a multifaceted process that includes all instructors and involves colleagues with expertise both in the subject matter and in standards of content and achievement in the field. This was viewed as a potential means of reducing the invidious correlation between grades given in the course and the scores given to the instructor by students.

Many commenters had no clear recommendation as to whether students should fill out evaluations anonymously, but some suggested that anonymous responders be tracked to ensure that no one filled out an evaluation more than once. While some commenters expressed support for grade quotas as an antidote to grade inflation, the large majority opposed it; outcomes assessment and the publication of average grades given in courses (which could be used by students to choose instructors with higher averages) were similarly opposed. Commenters instead recommended various other measures as counterweights to student evaluations, noting that such counterweights are particularly important when evaluations are included in tenure and promotion decisions. Peer review, coupled with the collective establishment of course content and grading norms by the faculty of the department or field, could serve as one such measure. For professional schools, the correlation between board exams and course grades could be helpful. Ideally, these measures would be combined with periodic meetings of the instructors of multisection courses or related courses to discuss common issues and share solutions and approaches. Such meetings could also provide the opportunity for another recommended approach, mutual mentoring by faculty members. An additional frequent recommendation was the development by faculty members of a teaching portfolio with a range of materials showing the standards set.

Respondents recommended that the evaluation instrument be developed by the faculty and that the questions be appropriate to the field and the pedagogical methods used. The most frequent recommendation was that the questions be carefully worded to avoid biasing student responses and that they focus on student learning: for example, the question should not be, "Did the instructor return work in a timely manner?" but rather, "Did the instructor return work before the results were to be applied to a later assignment?" Numerous commenters recommended a reflective component. Others reported adding their own evaluation forms, sometimes at midterm and frequently oriented toward the effectiveness of the course in promoting student learning.

Commenters recommended that students who have dropped a course or been charged with academic misconduct be excluded from evaluating the instructor. Some noted that a virtue of paper evaluations is that they are more likely to include responses from students who regularly attend the class—and less likely to include uninformed evaluations from students who are frequently absent.

A frequent recommendation was that course evaluations be treated as "a faculty development project"—that is, as formative rather than summative. They should be interpreted in the context of the course, and interpretation should take into account research on student evaluations and potential biases. Evaluations should not be used to impose conformity. Evaluators can gain perspective on an individual course by reviewing multiple courses taught by the instructor over multiple semesters, by reviewing the performance of the students of that course in subsequent related courses, and by including the instructor's former students among evaluators. Cross-evaluation of students by other faculty members can also be helpful.

Administrators should respect standards developed by the faculty and not exert pressure to obtain higher evaluation scores by lowering standards to please students, a problem frequently cited by commenters. Many commenters also observed that institutions can do much to protect educational standards by offering higher salaries and the protection of tenure.

Finally, respondents to the survey preferred that institutions support instructors' participation in field-based pedagogical conferences and

workshops, as well as research on methods, rather than fund generic teaching centers. They almost universally recommended pedagogical training in graduate programs.

### Additional Recommendations

The portrait that emerges from the survey suggests a number of additional recommendations.

First, student evaluations should be completed in class. The move to online evaluations completed outside of class appears to compromise whatever reliability one could hope to claim for student evaluations.

Additionally, faculty members within departments and colleges—not administrators—should develop instruments and determine practices (peer review, classroom visits, teaching portfolios) that reflect the kinds of courses being taught, the levels of the students in the courses, and the styles of teaching being promoted. University-wide or college-wide evaluation forms that disregard this variety should be avoided; they generate meaningless numerical comparisons that invite misuse.

Anonymity in student comments is necessary but may work against the gathering of reliable information by allowing students to make unfounded claims. Perhaps completing evaluations in real time in the classroom, though still anonymously, would curb this trend.

Fairness demands that changes be made in how institutions support teaching. Graduate students and non-tenure-track faculty members should be given access to the same teaching development opportunities offered to tenure-track faculty members. Moreover, chairs, deans, provosts, and institutions as a whole should not allow numerical rankings from student evaluations to serve as the only or the primary indicator of teaching quality. Publishing the results of student evaluations, as is done at various commercial rating websites, is counterproductive. The purpose of evaluation should be to help faculty members improve as teachers and to provide quality control; it should not be to help students find easy classes and avoid challenging teachers.

Perhaps most important, every department chair, dean, and provost should familiarize himself or herself with the AAUP's *Statement on Teaching Evaluation*. If that statement's policy recommendations were implemented more widely, many of the problems uncovered in the survey likely would not exist.

### Acknowledgments

The authors thank Martin Kich, Ann McGlashon, and Susan Michalczyk, the other members of the Committee on Teaching, Research, and Publication at the time when the survey was conducted, for their contributions to the survey and this article.

*Craig Vasey is professor of philosophy and chair of the Department of Classics, Philosophy, and Religion at the University of Mary Washington. He is a former member of the AAUP's national Council and current chair of the Committee on Teaching, Research, and Publication. His e-mail address is [cvasey@umw.edu](mailto:cvasey@umw.edu). Linda Carroll is professor of Italian at Tulane University and a member of the AAUP's Executive Committee.*

Tags: [Craig Vasey](#) [Linda Carroll](#) [Evaluation Of Faculty](#)

· [The Meaning of MOOC-topia](#)

[up](#)

[What Shakespeare Says about Sending Our Children Off to College](#) ·

Chris Nagel (not verified)

Fri, 05/27/2016 - 5:44pm

[permalink](#)

#### costs and burdens vs benefits

One issue this report does not get into is the proportionality of costs and other burdens to the benefits of good feedback on instruction. This is an issue for all, but especially for non-tenure-track faculty who most often work with limited or no access to faculty development resources, and whose livelihoods depend on good evaluations (and especially on high marks in student surveys). While it makes sense that informed and informative evaluation of instruction would benefit from a teaching portfolio, for instance, the cost and burden of producing one may be disproportionate for many term-contract faculty. I have struggled on my campus with trying to negotiate on this with administration (and in some cases with department chairs, who are not admin in our system). I have a strong union behind me, and serve as a union rep in those contexts. Individual non-tenure-track faculty are basically powerless to resist excessively burdensome or unfairly, unevenly applied procedures or criteria, even when there are terms of evaluation bargained for collectively. It's important to be mindful of these issues when considering recommendations for evaluation policies and procedures.

[reply](#)

lakins1

Sun,  
06/05/2016  
- 12:22pm  
[permalink](#)[cost-benefit unclear](#)

I too was thinking that there was little (or no) comment about the cost of the recommendations versus the benefits. Valuable and useful evaluation is time- and thus cost-consuming. This is true in every sector and from every "evaluator". By the way, I loathe calling the students' input as evaluative. It is an opinion survey, not much different from any consumer survey, and should not be raised to the level of evaluation of work which requires a discipline-specific knowledgeable evaluator. But back to the cost point ... Administrations seem to be willing to spend the money on administrative oversight instead of investing the money in the workers themselves. What evidence is there that the money is well spent? Why should we even bother spending the money on evaluation when the majority of faculty are working with no protections, no paid health insurance, and at a poverty wage?

[reply](#)

cguthrie

Sat,  
07/23/2016  
- 10:03am  
[permalink](#)[High Grades = High Evals](#)

As a relatively new department chair I now see student evals (done online) of a lot of faculty, both full and part-time. The only real information I gather from this are:

1. Evaluation averages correlate closely with the average grades that faculty give. I see this in both full and part time faculty. Given that we are a non-tenure teaching-oriented school in which evaluations play an outsized role in contract renewals, the pressure to grade high to get good evals is very strong. Of course, this gives us no information on whether a teacher is educationally effective or keeping collegiate standards high.
2. Student comments are generally only useful if the same factual comment is made several times. For example, I had a normally strong teacher receive evaluations in which several students complained about the teacher regularly shortening class times. This prompted a conversation with the teacher who confessed to having a semester over-stressed with personal challenges. It allowed us to talk about what to do if similar situations were to arise in the future.
3. The longest comments are invariably made by passive-aggressive students who are furious for some reason and didn't have the courage to talk directly with the faculty member or administrators. One student wrote a thousand outrageously angry words about being triggered by a comment by a professor that she took the wrong way. He was soon tendering his resignation out of humiliation. I've been excoriated in reviews by students who were clearly caught plagiarizing -- because, as they put it, no other professor had ever enforced those rules. This is simply Rate-My-Professor-level reflection. Students have little idea that their angry online words matter to the livelihoods of faculty who have committed themselves to helping students.
4. With faculty for whom I have several semesters worth of comments, it is useful to look at long-term trends in numbers. Along with narrative accounts of a teacher's efforts to improve, long-term trends can fill out a story about a teacher's growth in effectiveness (or not) over time.

I've come to believe that given the toxicity of the anonymous review culture in which we live, anonymous student evaluations are no longer very helpful.

[reply](#)