

Making the Grade: Using Instructional Feedback and Evaluation to Inspire Evidence-Based Teaching

Peggy Brickman,^{†*} Cara Gormally,[‡] and Amedee Marchand Martella[§]

[†]Department of Plant Biology, University of Georgia, Athens, GA 30602; [‡]Department of Science, Technology, and Mathematics, Gallaudet University, Washington, DC 20002; [§]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309

ABSTRACT

Typically, faculty receive feedback about teaching via two mechanisms: end-of-semester student evaluations and peer observation. However, instructors require more sustained encouragement and constructive feedback when implementing evidence-based teaching practices. Our study goal was to characterize the landscape of current instructional-feedback practices in biology and uncover faculty perceptions about these practices. Findings from a national survey of 400 college biology faculty reveal an overwhelming dissatisfaction with student evaluations, regardless of self-reported teaching practices, institution type, or position. Faculty view peer evaluations as most valuable, but less than half of faculty at doctoral-granting institutions report participating in peer evaluation. When peer evaluations are performed, they are more supportive of evidence-based teaching than student evaluations. Our findings reveal a large, unmet desire for greater guidance and assessment data to inform pedagogical decision making. Informed by these findings, we discuss alternate faculty-vetted feedback strategies for providing formative instructional feedback.

INTRODUCTION

Science faculty require a greater level of sustained encouragement and constructive feedback when implementing the evidence-based teaching practices that are currently advocated by groups such as the American Association for the Advancement of Science (AAAS) and the National Academies of Sciences (Henderson and Dancy, 2009; Henderson *et al.*, 2014; Gormally *et al.*, 2014). Yet for most faculty, end-of-semester student evaluations (Keig, 2000; Loehrer, 2006) and occasional peer observations (Seldin, 1999) represent their sole form of instructional feedback. Both mechanisms have considerable limitations (Gormally *et al.*, 2014, and references therein). For example, end-of-semester teaching evaluations recognize and reward student perceptions of teacher-centered behaviors rather than factors related to learner-centered pedagogies (Murray, 1983; Cashin, 1990; Marsh and Roche, 1993). These evaluations focus on student satisfaction rather than learning (Aleamoni, 1999; Kember *et al.*, 2002). As a result, instructors who incorporate learner-centered pedagogies in their courses may see declines in their student evaluation scores (Walker *et al.*, 2008; Brickman *et al.*, 2009; White *et al.*, 2010) or face student resistance through evaluation comments that may reinforce teacher-centered behaviors (Gormally *et al.*, 2016). In terms of peer evaluations, faculty have expressed lack of confidence in their peers' expertise (Bell and Mladenovic, 2008) and objectivity and concern over time constraints and potential detrimental career implications (Centra, 1993; Atwood *et al.*, 2000; Chism, 2007; Kell and Annetts, 2009). It is difficult to even estimate the number of faculty participating in peer instructional evaluation, the type of feedback they receive, or their perceptions of the value of peer evaluation in general. Consequently, there is a need to characterize the current state of feedback practices to determine whether these practices hinder the effective adoption and persistent use of evidence-based teaching practices.

Diane K. O'Dowd, *Monitoring Editor*

Submitted December 3, 2015; Revised April 13, 2016; Accepted April 14, 2016

CBE Life Sci Educ December 1, 2016 15:ar75

DOI: 10.1187/cbe.15-12-0249

*Address correspondence to: Peggy Brickman (brickman@uga.edu).

© 2016 P. Brickman *et al.* CBE—Life Sciences Education © 2016 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.

Yet peer-feedback processes do have demonstrated potential to improve teaching practices (Weimer, 2002; Pershing, 2006; Stes *et al.*, 2010; Gormally *et al.*, 2014). Feedback from faculty peers can provide evidence of performance on aspects of teaching such as depth of subject knowledge and appropriateness of course material that are better assessed by peers than students (Peel, 2005; Berstein, 2008). Positive outcomes of peer feedback can include improved self-assurance (Bell and Mladenovic, 2008), collegiality and respect (Quinlan and Bernstein, 1996), and improved classroom performance (Freiberg, 1987). However, it is important to distinguish between the types of peer feedback faculty may experience.

Peer feedback may be either formal (and evaluative) or formative. Formal peer review and evaluation involves observing and measuring teaching for the purpose of judging and determining “value.” Formal peer review and evaluation is usually part of merit, promotion, and/or tenure decisions. Instructors are not likely to seek candid feedback if they may be penalized for doing so on formal evaluations (Cosh, 1998; Martin and Double, 1998; Shortland, 2010). Formative peer feedback, on the other hand, is meant to assist in the improvement of teaching. This process is designed to be informal, supportive, sympathetic, and friendly, with frequent contact between the parties involved, as summarized by Gormally *et al.* (2014). A peer-feedback model that is collaborative and formative in nature, rather than one-sided and evaluative, seems more likely to transform faculty teaching practices.

To date, there has been no systematic analysis of the current state of instructional feedback given to college biology faculty or work to uncover what unmet needs may exist or how faculty use feedback to improve their courses. The goal of this study was to characterize the landscape of current instructional-feedback practices in biology departments. Additionally, this study sought to uncover unmet faculty needs and wishes for instructional feedback and resources that might support more effective implementation of evidence-based teaching practices. We were interested in understanding what types of incentives might motivate faculty to participate in giving or receiving formative instructional feedback, what ego-saving devices would need to be in place, and who faculty would respect as feedback givers. Finally, we wanted to learn about areas in which faculty sought feedback: in-class teaching performance, curricular materials, or course design and structure. Specifically, our study addressed the following research questions:

1. What are the current instructional-feedback practices in college biology programs? Do these current practices reflect what we know about best practices for feedback?
2. What are faculty’s perceptions about current practices?
3. What do faculty say they need in terms of instructional feedback?

METHODS

Survey Development

We developed survey questions designed to collect faculty perceptions about current instructional-feedback practices and elicit faculty wishes for ideal feedback practices (see the Supplemental Material). We also collected faculty demographic data so we could compare responses to these items between faculty at different institutions, of different ranks, and with dif-

ferent experiences. We crafted these questions to incorporate what is currently known about effectiveness of feedback practices from Organization Psychology and teacher practices in the K–12 setting (Gormally *et al.*, 2014) and research characterizing the current reality of instructional-feedback practices. We then conducted individual cognitive interviews using a think-aloud protocol (Ericsson and Simon, 1980; Collins, 2003) with initial survey questions. Twelve faculty members were asked to respond to each survey question and, as they answered the question, to describe why they responded in particular ways, to explain whether item responses were missing or irrelevant, and to comment on question and item-response clarity. We selected a range of faculty to represent different tenure status and institution type. After each interview, two coauthors (P.B. and A.M.M.) discussed and revised relevant survey items to address ambiguous wording and add or revise items.

Sampling Methodology

We randomly selected a subset of institutions from each category in the Carnegie Classification of Institutions of Higher Education (McCormick, 2015). This classification includes both public and private institutions that are doctoral, master’s, baccalaureate, and associate’s granting (for definitions of each category, see McCormick, 2015). Using website research, we obtained email addresses from up to six faculty from the biology departments of these institutions, attempting to get an even distribution of faculty from different ranks (two full professors, two associate professors, and two assistant professors or lecturers). In the case in which a randomly selected bachelor-granting or doctorate-granting institution had multiple departments in the biological sciences but no formal biology department, we selected up to six faculty using the same distribution from the various departments at these institutions (e.g., organismal biology, molecular and cellular biology, evolutionary biology and ecology, environmental science). If a randomly selected institution defined its academic departments more broadly than biology (e.g., “life sciences” or “sciences”), we obtained email addresses for those faculty with biology backgrounds rather than those who had other science backgrounds (e.g., chemistry). Institutions were omitted for the following reasons: 1) biology-specific directories for faculty could not be found on the website; 2) faculty emails were not listed, or only online email boxes from which to send questions were provided (this was uniformly true at for-profit institutions, which we did not survey); 3) biology or other relevant science departments did not exist at the institution; or 4) some institutions and categories of institutions were too specific (e.g., Le Cordon Bleu College of Culinary Arts).

Using this methodology, we obtained email addresses for faculty at 70 doctoral institutions (24% of all 297 doctoral institutions), 175 master’s institutions (24% of 724 master’s institutions), 161 baccalaureate institutions (24% of 663 appropriate baccalaureate institutions), and 344 associate’s institutions (33% of 1042 associate’s institutions). We sent the URL for our Web-based survey of feedback practices to more than 4000 faculty in October 2014 (Qualtrics, 2016). We received a total of 399 individual responses, of which 343 were complete for all questions. The respondent pool represents at least 185 different institutions (102 of the respondents did not include the names of their institutions). We checked the IP addresses of responses from the same institution to be sure that responses were unique

to ensure that we did not receive more than one response from any individual.

Data Analysis

Our survey included categorical response questions with both ordinal categories, such as class sizes (20 or fewer, 21–50, 51–200, and more than 200 students) or satisfaction with end-of-course student evaluations (“yes,” “in some ways,” and “no”), and nominal categories, such as type of institution (associate’s, baccalaureate, master’s, and doctoral granting), Likert-scale questions, and open-ended questions. We used chi-square (χ^2) analyses to analyze categorical response survey questions. When expected values were less than 5, Fisher’s exact test was used. We also used regression analysis to investigate relationships between multiple variables. Ordinal logistic regression was conducted for categories that contained answers with natural ordering, such as “yes,” “in some ways,” and “no.” For demographic questions, we report descriptive statistics to show both the range of responses and means, where appropriate. We coded the responses to open-ended survey questions with our research questions in mind:

1. What are the current instructional-feedback practices in college biology programs? Do these current practices reflect what we know about best practices for feedback?
2. What are faculty’s perceptions about current practices?
3. What do faculty say they need in terms of instructional feedback?

We developed a unique codebook for each open-ended survey question, because we viewed each question as an individual data set. For each question, one researcher coded initially to begin our coding process. Then, the coauthors used the initial codebook to independently code the responses, while adding, modifying, and clarifying codes as needed. Finally, we reviewed the coding together, discussing and merging codes through consensus and adding emergent codes as needed.

Demographics of Survey Respondents

We emailed our survey to more than 4000 faculty, and received 399 responses (Table 1). Despite efforts to contact multiple faculty at each institution, 82% of respondents were the sole respondents at their institutions. One institution (in the doctoral category) had five respondents; however, this represents only 1.3% of the total sample and only 6.9% of the responses in the doctoral category. Our responses are similar to faculty demographic data from the National Center for Education Statistics (NCES, 2011). In our data set, professors may be slightly overrepresented (33.3% vs. 23.8% in NCES data). We received more responses from women than men; consequently, women are overrepresented in our data set (55%; NCES indicates women represented 44% of faculty in 2011). Nationally, 78.4% of natural sciences faculty are white (NCES); our data set shows a slightly lower percent of faculty who identify as white or Caucasian; however, we had a significant number who did not report. We asked survey participants to report their position title; the data were consistent with NCES statistics. We report 23% full; 20% associate; 23% assistant; 18% instructors/lecturers; 14% other (Table 1). The majority of respondents did not currently serve in administrative roles (64%); however, 23% of respondents were unit heads (department, division, or program

TABLE 1. Respondent demographics

Respondents	N	Percentage of total
Gender		
Female	190	47.6
Male	142	35.6
Prefer not to respond	11	2.8
No response	56	14.0
Race		
White	296	74.2
Asian	10	2.5
African American	7	1.8
American Indian or Alaskan Native	2	0.5
Other	11	2.8
Prefer not to respond	26	6.5
No response	47	11.8
Ethnicity		
Hispanic or Latino/a	6	1.5
Not Hispanic or Latino/a	284	71.2
Prefer not to respond	33	8.3
No response	76	19.0
Rank		
Professor	132	33.2
Associate professor	94	23.6
Assistant professor	90	22.6
Instructor	43	10.8
Senior lecturer	5	1.3
Lecturer	6	1.5
Adjunct	15	3.8
Other	8	2.0
No response	6	1.5
Type of institution		
Doctoral granting	72	18.0
Master’s	59	14.8
Baccalaureate	148	37.1
Associate’s	113	28.3
No response	7	1.8

heads). Nine percent of respondents reported they were teaching mentors; it is important to note that we left this open to interpretation by the participants, but we assume this means they provide support either officially or otherwise for other faculty to improve their teaching.

Our respondents reported teaching for 15.8 years (after graduate school) on average, with 65.8% having taught for 10 or more years, and 18.1% having taught for five or fewer years. We asked faculty to report their average teaching responsibilities and course enrollments (Table 2). As expected, there is a difference between number of courses taught and institution type. (A Fisher exact test performed after combining classes into three categories—“less than 2,” “2 or 3,” and “greater than 3”—resulted in a p value of < 0.0001 .) Respondents at associate’s-, baccalaureate-, and master’s-granting institutions tend to teach more than four classes in a typical year, while respondents from doctoral-granting institutions most commonly reported (35%) teaching two classes in a typical year. Only 15% of respondents from doctoral institutions taught more than four classes in a typical year, versus 82% of associate’s, 78% of baccalaureate, and 62% of master’s respondents.

TABLE 2. Respondent class sizes and number of classes taught

Respondents	N	Percentage of total
Number of classes taught		
0	2	0.5
1	16	4.0
2	36	9.0
3	27	6.8
4	54	13.5
>4	255	63.9
No response	9	2.3
Class enrollment		
Under 20	87	21.8
21–50	221	55.4
51–100	41	10.3
101–200	22	5.5
200–400	13	3.3
> 400	4	1.0
Not applicable	2	0.5
No response	9	2.3

Our survey also sought to uncover whether biology faculty use evidence-based teaching practices such as active learning in their classes. Nearly half (48.3%) of respondents report that they lecture 50% or less of the time during a typical week of class. In fact, 13.8% report that they lecture 25% or less of the time. However, 27.9% of faculty report lecturing 75% of the time or more, but only 4% of respondents report lecturing 100% of the time. During a typical week, faculty report spending on average 56% of class time lecturing, 24% of the time supporting students as they work in groups, 12% of the time facilitating whole-class discussions, and ~9% of the time supporting students working individually.

Respondents typically had few pedagogical development experiences before their faculty positions (Table 3), which aligns with other reports that graduate training does not provide extensive teacher preparation (Austin, 2002; Richlin and Essington, 2004; Foote, 2010). Only 29% of respondents reported having attended a new teaching assistant orientation training, and 18% reported attending other graduate teacher training programs such as a future faculty program. Respondents' experiences with professional development were consistent with other studies, with 76% reporting having attended workshops and seminars sponsored by centers for teaching and learning or professional societies (Henderson and Dancy, 2008; Fairweather, 2009). Respondents reported significant interactions with their faculty peers, with 70% of respondents reporting they had informally observed another faculty member teaching a course, 63% reporting they had received a peer observation, and 32% reporting they had served as a faculty facilitator/trainer at workshops. Nearly 40% (39%) reported receiving mentorship in teaching as a faculty member. The vast majority of respondents (97%) reported taking advantage of professional development opportunities.

RESULTS

Here, we report responses for survey items. Percentages are calculated based on responses given; those respondents giving no answer are not included.

TABLE 3. Respondent participation in professional development

Respondents	N	Percentage of total
Participated in teaching workshops or seminars (e.g., at centers for teaching and learning, professional societies) as a faculty member	287	76
Informally observed someone else teaching a course	263	70
Received peer faculty teaching evaluations (when a peer or supervisor attends one or more class sessions and creates a written or oral report)	235	63
Conducted peer faculty teaching observations	193	51
Served as a teaching mentor for a peer faculty member	163	43
Received mentorship in teaching as a faculty member	146	39
Served as a faculty facilitator/trainer at workshop or program	119	32
New teaching assistant orientation training	110	29
Graduate teacher training programs (future faculty program, teaching certificate)	67	18
I have not participated in any professional development	13	3

Faculty Perceptions about Current Instructional-Feedback Practices

End-of-course student evaluations were used by the vast majority of participants' institutions to evaluate the quality of both junior faculty (97%; including adjuncts, lecturers, and assistant professors) and senior faculty (90%). Yet faculty were not satisfied with their current official end-of-course student evaluations (Figure 1A). Forty-one percent reported that they were dissatisfied, while 46% reported that they were satisfied "in some ways." Faculty cited many reasons why they were dissatisfied with student evaluations (Figure 1C). Their top five reasons included: evaluations do not provide constructive information; poor response rates; the evaluations do not align with instructors' objectives; the evaluations only measure student satisfaction; and the process is not set up to truly engage students to attain useful and insightful feedback.

We had predicted that feedback practices might differ at baccalaureate-granting colleges, where a higher value is traditionally attached to teaching. However, our data show no evidence to suggest that this is the case in terms of end-of-semester student evaluations. Faculty at all types of institutions reported similar low levels of satisfaction with official end-of-course evaluations (Fisher's exact test, $p = 0.1004$). Interestingly, faculty teaching classes with enrollments of 50 or fewer students report using their student evaluations more to make improvements than faculty teaching in large-enrollment courses (Figure 1B; Fisher's exact test, $p = 0.0139$). Perhaps most tellingly, our regression analysis showed no relationship between satisfaction with official end-of-semester student evaluations and self-reported teaching practices (amount of time spent lecturing vs. using active-learning strategies; $\chi^2(1) = 0.7680$, $p = 0.3808$), institution type ($\chi^2(3) = 0.7923$, $p = 0.8513$), position ($\chi^2(5) = 2.7289$, $p = 0.7417$), years teaching ($\chi^2(5) = 5.6373$, $p = 0.3431$), and involvement in professional development related to teaching ($\chi^2(1) = 2.8530$, $p = 0.0912$).

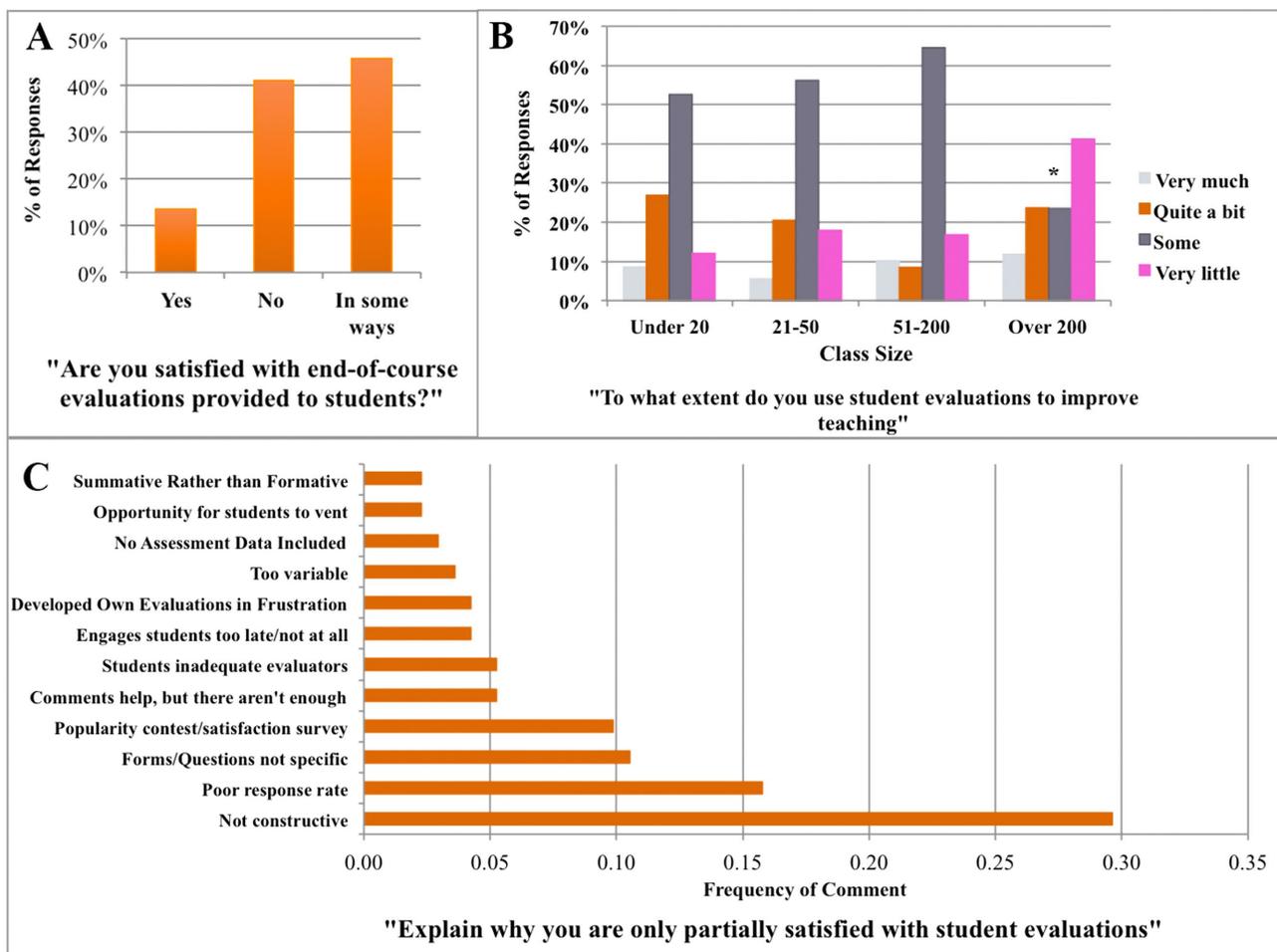


FIGURE 1. Faculty satisfaction with end-of-semester student evaluations. Survey participants indicated (A) their degree of satisfaction with end-of-semester student evaluations, (B) the degree to which they used student evaluations to improve their courses, and (C) the reason for dissatisfaction with student evaluations. The 40% of faculty teaching classes with more than 200 students reported using evaluations less than faculty teaching smaller classes. Asterisks indicate Fisher exact test p values < 0.05 . For B, the most frequent responses ($N = 304$) are indicated. (Only responses given more than five times are shown.)

On the basis of prior research (Bell and Mladenovic, 2008; Kell and Annetts, 2009; Thomas *et al.*, 2014), we anticipated that peer-teaching observations would be commonly performed, and indeed, 69% of survey respondents reported that their departments conducted peer-teaching observations using either fellow faculty members or departmental administrators (Figure 2A). However, our data suggest that peer observations are conducted more often at teaching-focused institutions (77–72% faculty at associate's-, bachelor's-, and master's-granting institutions, Fisher's exact test, $p = 0.0005$, Figure 2B) compared with doctoral-granting institutions. At doctoral-granting institutions, only 44% of faculty reported engaging in peer-teaching observations. When asked to rank the different methods available for evaluating teaching, peer evaluations were most valued by respondents (Figure 2C). Although these evaluations were official and evaluative, only 53% of respondents indicated that their department used a standard form for conducting peer-teaching observations. Faculty were also asked to indicate the criteria used to choose evaluators. The most common reason given was administrative responsibility (such as department head or assigned teaching

evaluator, 47% of 206 responses) and availability of colleagues and in the same discipline (43% of 206 responses, respectively). Faculty evaluators were also selected for years of teaching experience (22%) or seniority (27%) but only rarely for having received teaching awards (1%) or garnering high teaching evaluations (6%). Interestingly, 22% of the 206 respondents indicated novel criteria used to select evaluators, such as a shared system in which all faculty participate, volunteers, peer committees, junior faculty interested in observing, assigned mentors, or personally selected observers. Faculty typically receive feedback about peer evaluations in a timely manner, with 47% of faculty waiting for less than 1 week and 87% receiving feedback within the same semester.

Sixty percent of the participants responded to a question that asked about receiving peer comments during their peer-teaching observation, with 95% of these reporting receiving comments. It is important to note that only 69% of faculty reported that their departments conduct peer-teaching observations. When asked to select categories related to most frequently received comments, faculty reported most

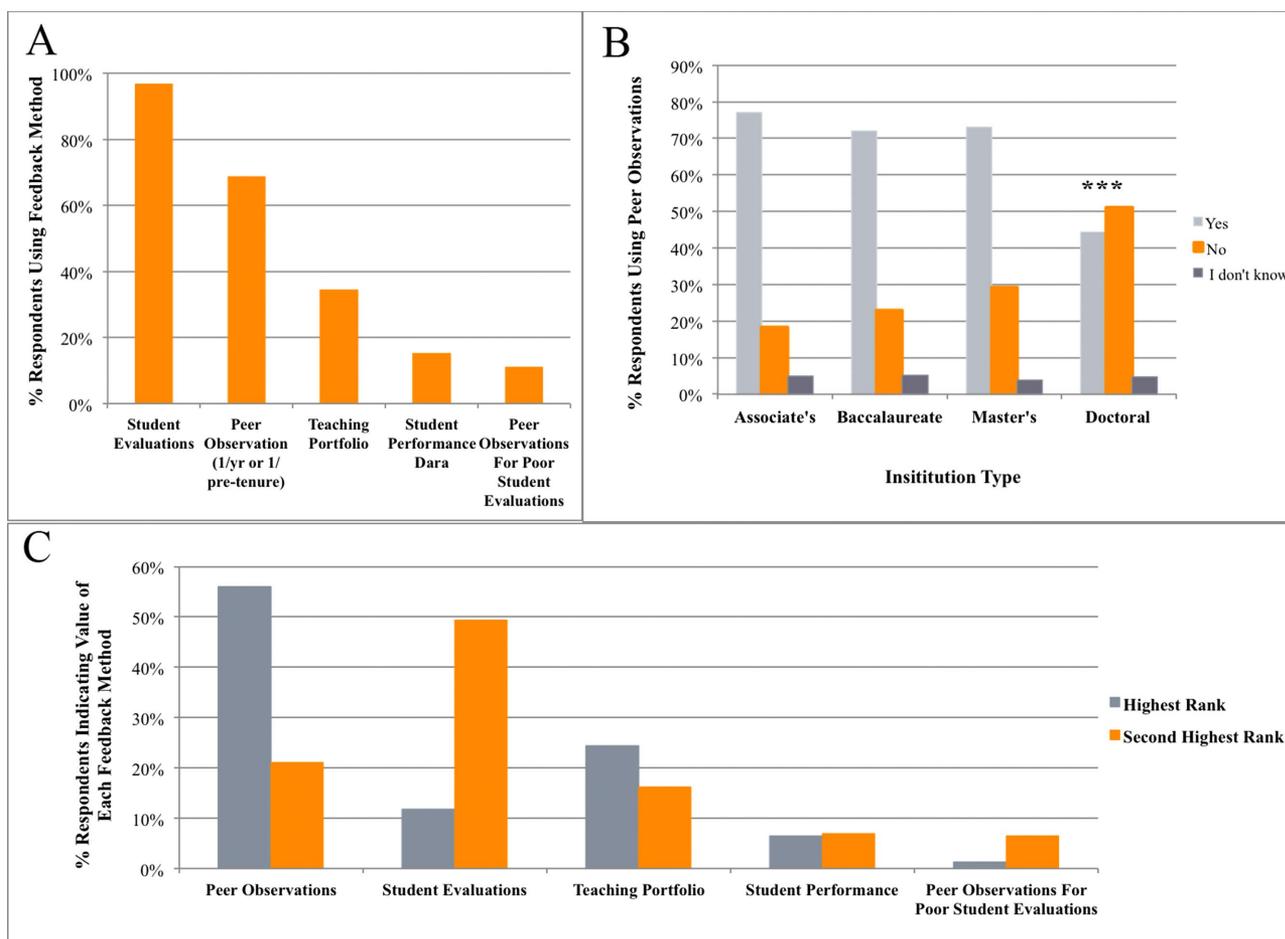


FIGURE 2. Common instructional-feedback methods used. Survey participants indicated which of the most common types of instructional-feedback methods were currently used in their department (A). When comparing different institution types for use of peer observations of teaching (B), the likelihood of conducting peer observations was significantly lower at doctoral-granting institutions compared with all other types (Fisher’s exact test, $p = 0.0005$). Respondents were also asked to consider the value of each of the different methods of providing instructional feedback used at their institution for both junior and senior faculty and rank the methods in order of value from highest ranking to lowest. Frequencies of the responses for highest and second-highest ranked methods are shown for each category for both junior and senior faculty combined (C). Counts were summed for each feedback method for junior and senior faculty, since chi-square tests of independence showed no significant differences in the counts for highest and second-highest rankings of each feedback method ($p > 0.05$).

frequently reported comments concerning rapport and interactions with students (69 and 73%) and feedback about more lecture-related behaviors such as providing clear explanations (65%), organization (60%), speaking style (56%), content (50%), or demeanor (46%). Faculty less commonly received comments about: how learning was managed, including time management (38%); providing learning objectives (28%) or class goals (25%); effectiveness of class activities (43%); and quality of assignments and tests (15%).

We sought to understand what types of changes faculty made to their teaching practices as a result of receiving instructional feedback via peer-teaching observations. Respondents were asked, in two open-ended questions, to first describe specific comments they have received in peer observations and subsequently to report any changes they had made as a result of this feedback. The most common response that faculty gave to the first question ($N = 166$) revolved around the comments

received during feedback that were not constructive and did not focus on student learning (30% of all comments; Table 4). Comments were coded as “not constructive” when they indicated a lack of focus on student learning, when suggestions for improvement were limited or vague, and when responses indicated the process was a formality rather than a real opportunity for growth. Other frequently reported comments were related to rapport (20%) and instructor behaviors such as PowerPoint presentations or use of the board (14%).

The type of feedback that faculty received had a substantial impact on their responses and whether they made changes to their teaching. For example, only a minority (23%) of faculty who did not receive constructive feedback reported making changes to how they were teaching. Sixty-seven percent of faculty receiving other types of specific feedback reported making changes in response. For example, faculty most frequently reported changes that focused on instructor behaviors related

TABLE 4. Unconstructive comments faculty received via faculty peer observations

- They commented mostly on presentation with little comment on the learning that they observed or the appropriateness of the material taught.
- Often the comments I have received are simply a list of topics we addressed in class and whether I was engaged with the students. There hasn't been much constructive criticism that is helpful. However, we have other faculty who evaluate junior faculty who do a much better job at classroom visit reports.
- My feedback was in the form of a two-page letter (to be included in my pretenure review file) that highlighted what occurred during the class observed, what I did well, and one suggestion for improvement.
- Comments were generally positive with explanations as to why they offered positive feedback. No suggestions for change.
- No specific helpful comments. Just one of those "you did great" comments when I know that there were things that I could improve. Also, there was only a one time observation.

to presenting class content, such as modifying their presentations and presentation style (16%). Interestingly, nine of 12 faculty ($N = 166$ total) who were encouraged to try active learning in their peer observations reported subsequently implementing active-learning pedagogies.

We hypothesized that faculty were dissatisfied with current practices at least in part due to receiving conflicting messages. When we asked, we found that faculty largely agreed with their peers' comments on observations (60% agree or strongly agree). However, nearly half (49%) those surveyed indicated they had received conflicting comments between peer observations and student evaluations. This conflict was consistent for all categories of faculty ranks ($\chi^2, p = 0.50$) and institution type ($\chi^2, p = 0.46$). When asked to provide further explanation of these discrepancies, faculty reported that their overall faculty peer observations were more positive than student evaluations (22% of 68 applicable coded comments). More specifically, respondents commented that, compared with faculty peer observers, students found the course too rigorous (24% of applicable coded comments); students did not appreciate teaching strategies (16%), often complaining about active learning; and students felt the course lacked clarity (10%) and organization (4%), and identified problems with the instructors' accent or speaking style (4%; Figure 3). In only 3% of coded applicable comments did faculty indicate that student evaluations were more positive than faculty observations. These responses illuminated the fact that faculty peers have a very different and more positive perspective about their peers' teaching practices than students. We hypothesized that faculty peer observations might be hampering adoption of active-learning practices; however, in reality, many respondents described that faculty peer observations supported the adoption of these practices.

What Do Faculty Say They Need in Terms of Instructional Feedback?

When asked "What's missing now when you seek feedback about teaching?," faculty focused mostly on increasing and improving opportunities for constructive peer observation (Figure 4A). However, respondents noted that the many demands on faculty time limit opportunities for peer observation

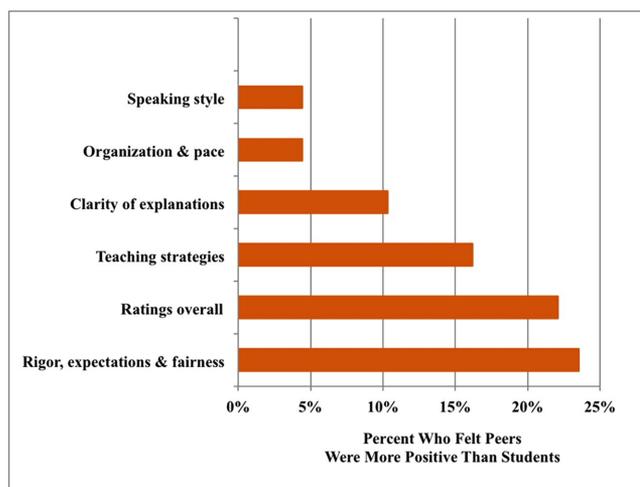


FIGURE 3. Reconciling student evaluations and faculty peer observations. Survey participants ($N = 88$) provided open-ended responses to the question "Please describe your experience with conflicting comments from student evaluations and faculty peer observation in the space below." Almost a third of all comments were not applicable to the question, for example, when faculty provided explanations of conflicts among student evaluations but not between student evaluations and peer observations. We identified 16 different qualitative code categories out of the remaining comments. The top six codes are indicated with frequencies expressed as a percentage of applicable codes ($N = 68$), and all provide examples in which faculty felt peer evaluations were more supportive of their teaching than were student evaluations. Less frequently expressed themes included discrepancies on perceptions of instructor interest and enthusiasm, rapport with students, appropriate use of technology, gender bias, and the perception that faculty observers provide negative and positive comments, while students are more effusive but vary widely in the range of comments they provide.

and evaluation and that administrators do not reward or value teaching. Despite these constraints, 41% of respondents indicated they proactively seek feedback several times each semester to bridge the gap in instructional feedback, most commonly from colleagues (83%) and teaching workshops (67%; Figure 4B). Faculty also mentioned using professional organizations (44%), educational research literature (59%), and pre- and posttest assessment data (35%; Figure 4B). Faculty most commonly sought feedback about class activities (81%), classroom assessment techniques (61%), and presentation materials (e.g., PowerPoints, handouts; 54%). Perhaps unsurprisingly, 54% of respondents indicated they sought social support (e.g., empathy, reassurance, commiseration) as well. Twelve percent of respondents indicated that they never seek additional instructional feedback. When asked to indicate how often they would like to receive feedback, most faculty wanted to receive feedback once a year (20%), once a semester (35%), or two to three times each semester (31%). Fewer than 3% of respondents wanted to receive feedback more than five times a semester. Only 5% of respondents reported that they never wanted to receive feedback. Some faculty indicated that their answer depended on whether the feedback was formative or summative. One respondent stated, "Feedback 2–3 times a semester would be much

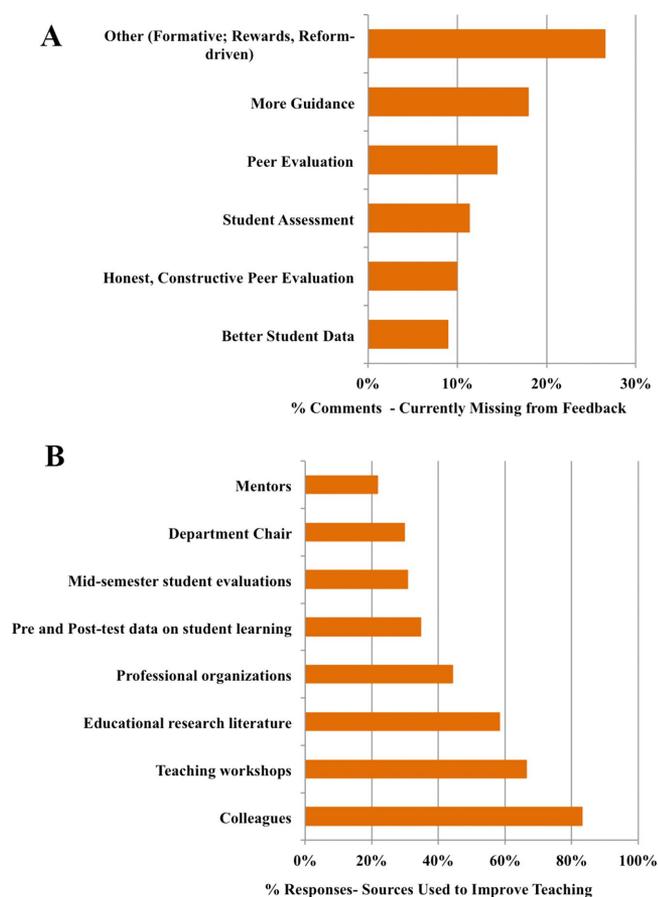


FIGURE 4. What is missing from instructional feedback? (A) Survey participants ($N = 235$) provided open-ended responses to the prompt “Describe what you think is currently missing from the feedback you are getting about your teaching.” We identified 15 different qualitative code categories. Nine percent of faculty were satisfied with the current system. The top six codes for answers to this question are indicated with frequencies expressed as a percentage of total codes ($N = 289$). Less frequently expressed themes (unpublished data) included requests for formative feedback, receiving feedback in a more timely manner, feedback offered through the process of team teaching or faculty learning community, feedback in tandem with comparative student assessment data, reform-based instructional feedback, formal feedback of any kind, rewards, demands, and comments that were not applicable to the question. (B) When asked to select from a list of sources faculty have voluntarily accessed to improve their teaching, survey participants ($N = 347$) selected these categories most frequently.

appreciated, but formal evaluation that will impact the tenure decision would be too stressful at that frequency.”

If the choice of feedback source were theirs alone, faculty continued to select both students and their peers as valuable sources of feedback, and this was true regardless of institution type ($\chi^2(24) = 30.45, p = 0.17$). Rather than relying on typical student evaluation ratings, faculty indicated a desire to turn to novel strategies such as midsemester student evaluations, data about student learning, and alumni evaluations. Faculty reported that they would seek voluntary peer classroom observations conducted by a faculty peer of their choosing (64%) or faculty with

education research experience (43%). Characteristics of ideal feedback providers include a faculty peer who is an experienced and even renowned teacher, a peer who teaches similar courses, and someone familiar with innovative evidence-based teaching strategies or knowledge of education research. From survey results, it is clear that faculty seek substantive, honest, and formative feedback. Respondents frequently commented that this faculty peer must provide constructive feedback and have well-developed interpersonal skills to conduct this type of dialogue, such as being approachable, supportive, adaptive, and sensitive to an individual faculty member’s needs and situation. Interestingly, respondents were less concerned about whether the faculty peer was tenured, was junior faculty, and came from inside or outside their departments. Faculty want to receive feedback from peer observations in a timely manner, either immediately after (12%), within 48 h (24%), or within a week (53%).

We asked faculty, “Please provide an example of an incentive that would entice you to seek feedback on your teaching.” The most frequent responses were monetary incentives (e.g., raise, bonus); the satisfaction of improving their students’ learning; the intrinsic reward of improving one’s teaching and the idea that “faculty should want to do this”; and the incentive of promotion, continued employment, and annual evaluations (Figure 5A).

Interestingly, faculty responses about types of incentives to motivate them to improve their teaching differed dramatically from the benefits they purport to accrue from actually making changes to their teaching practices. When asked about the benefits they have realized as a result of making changes to their teaching, faculty reported rewards in terms of improvements in student evaluations, rapport with students, and achievement at a much greater extent than rewards to themselves (awards, recognition, or monetary salary increases; Figure 5B).

Study Limitations

This study has a few limitations, including reliance on self-reported data, which is inherent to studies using survey data. We asked respondents to report their typical teaching practices; self-report data may be more subjective than other forms of data collection and survey questions may be interpreted differently by individual participants. To address the latter issue, we conducted cognitive interviews (described in the *Methods*) to refine and clarify the survey items. Given the survey’s length, it was not possible to include more survey items (e.g., items related to the Classroom Observation Protocol for Undergraduate STEM or other validated teaching practice inventories) to more rigorously evaluate reported teaching practices (Smith *et al.*, 2013; Thomas *et al.*, 2014). The survey response rate was relatively low, with 399 participants out of a potential pool of 4000. It is also possible that response bias influenced our findings, for example, faculty inherently more interested in teaching may have responded at a higher rate than faculty who dislike teaching.

DISCUSSION

This study provides a concrete characterization of the current instructional-feedback practices in college biology programs. Given faculty responses, we were able to evaluate whether these practices align with feedback best practices and characterize faculty perceptions and expressed needs for instructional

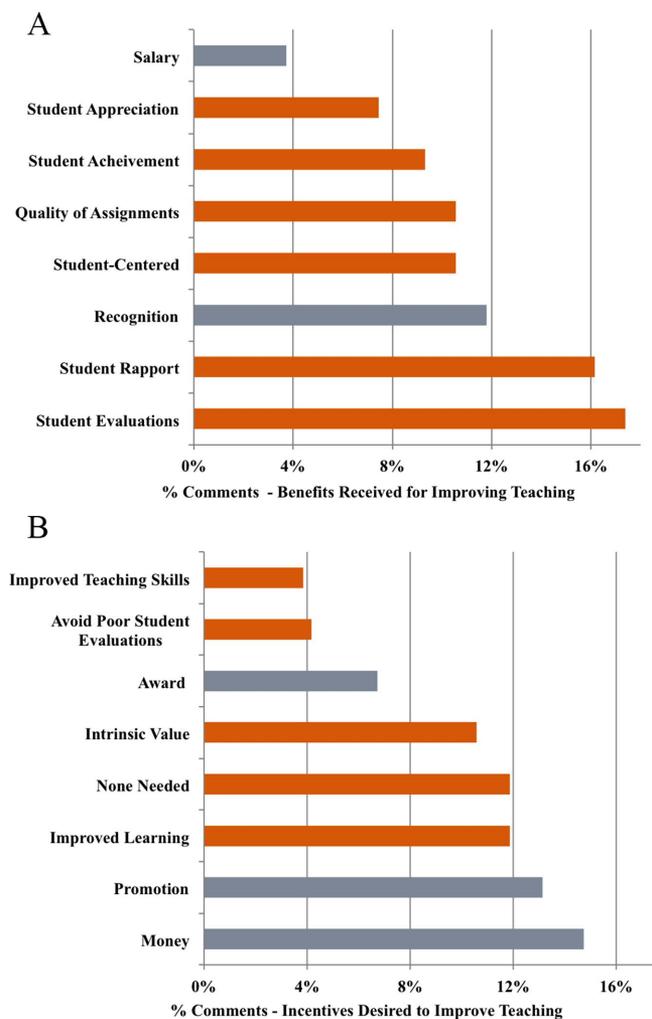


FIGURE 5. Benefits accrued from changing teaching practices. (A) A group of 147 survey participants provided 161 examples of specific benefits they received as a result of making changes to how they teach. When the categories of answers were analyzed, 15 different codes were identified. The eight most common codes are shown. Gray codes indicate direct extrinsic benefits to faculty, while orange codes indicate intrinsic benefits. **(B)** A group of 245 survey participants provided 312 examples of incentives that would entice them to make improvements to their teaching. When the categories of answers were analyzed, 24 different codes were identified. The eight most common codes are shown. Gray codes indicate direct extrinsic incentives to faculty, while orange codes indicate intrinsic incentives.

feedback. While our study focused on biology programs, we expect that other researchers may find that feedback practices are similar in other science, technology, engineering, and mathematics (STEM) fields. We summarize our major findings in the paragraph below. We then focus on applied outcomes that emerge from what faculty say motivates them to change. We consider how novel strategies mined from what faculty are doing to bridge this gap might be leveraged for new models of instructional feedback.

Current instructional-feedback practices are not providing the support that biology faculty seek. We found that faculty are

dissatisfied with the current student evaluation systems. In particular, faculty teaching large-enrollment classes reported much lower use of student evaluations to improve their classes. Yet faculty respondents continue to believe that students are their most important source of data, suggesting the need to reimagine the types of student data used to evaluate teaching. Student ratings of faculty will always play a critical role in illuminating perceptions of instruction. Student perceptions of instruction impact retention in the major; poor evaluations have been linked to an increased likelihood of a student dropping a class and a decreased likelihood of taking subsequent courses in the same subject (Hoffmann and Oreopoulos, 2009). Richardson (2005), Stark and Freishtat (2014), and Wachtel (1998) all present excellent reviews of studies examining the efficacy of student evaluations of teaching and document problems with perpetuating the use of existing student evaluation instruments, even those built upon research evidence with validity and reliability (such as the Student Evaluation of Educational Quality). Use of these evaluations is suggested to reinforce didactic models of teaching and reinforce students' expectations of didactic teaching that often run contrary to teachers' own perceived needs (Ballantyne *et al.*, 2000). Student evaluations of teaching often exhibit low response rates that introduce sampling error or bias. Research has shown that students who respond voluntarily to surveys are different from those that do not respond in terms of study behavior and academic attainment (Nielsen *et al.*, 1978; Watkins and Hattie, 1985). In addition, although Feldman (1993) reported that the majority of studies of student evaluations of teaching found no significant differences between the genders, several recent studies have indicated bias for gender and ethnicity of the instructor (Basow, 1995; Anderson and Miller, 1997; Cramer and Alexitch, 2000; Boring *et al.*, 2016). Acknowledging these problems may help convince colleagues to use objective measures for determining the effectiveness of courses like the teaching portfolios (collections of documents that include a variety of different types of information including evidence of student learning) that more than 30% of our respondents indicated using or measures of student learning (e.g., course products, progress between initial and summative work, or written reflections as suggested by Blumberg, 2014).

We found that faculty highly value peer observations. Yet, like student evaluations, peer observations are not without their share of problems. First, peer observations are not often conducted uniformly, and only about half of respondents reported the use of an official form or feedback template. Peer observations are irregularly conducted at doctoral institutions, and faculty responses suggest that, in some cases, peer observations may be a rubber stamp rather than a real opportunity for critical feedback. One major discrepancy faculty report was that peer observations tended to be more positive than student evaluations. In fact, respondents were concerned that their peers did not offer enough honest, constructive feedback. Yet faculty reported making changes to their teaching practices when they received rigorous feedback from their peers. Encouragingly, faculty expressed desire to revise the instructional-feedback process in order to receive more critical feedback.

Improving mechanisms for instructional feedback may be the missing link for more effective implementation of evidence-based teaching strategies in undergraduate STEM education. There

have been enormous efforts to promote instructional change nationwide (AAAS, 2010; Anderson *et al.*, 2011; President's Council of Advisors on Science and Technology, 2012; Association of American Universities, 2014; National Research Council, 2015). However, despite the efforts of change agents, instructors are not implementing evidence-based strategies as prescribed by developers (Turpen and Finkelstein, 2009, 2010), and modification of instructional practices has been associated with reduced learning gains in other studies (Andrews *et al.*, 2011; Henderson *et al.*, 2012; Chase *et al.*, 2013). Many instructors who attempt to use evidence-based strategies later quit, naming both student and peer resistance as factors (Henderson *et al.*, 2012). To support instructional change, faculty clearly need more than just knowledge of effective teaching strategies. They also need motivation, support, critical reflection, and concrete suggestions for improvement (Gormally *et al.*, 2014; Bradford and Miller, 2015; Ebert-May *et al.*, 2015).

Interestingly, in our study, faculty peers provided much greater support for evidence-based teaching practices than student evaluations. This may suggest that the majority of biology faculty now recognize that evidence-based strategies have proven potential to improve student learning (Freeman *et al.*, 2014). However, our faculty respondents reported that the limited amount of constructive feedback they do receive from their peers focuses on lecture presentation skills, with less emphasis on student learning and assessment. These findings suggest that faculty may not be offering constructive feedback about evidence-based teaching, because they may need more support to implement these strategies themselves. For example, few active-learning instructors have ever observed another instructor using active learning (Andrews and Lemons, 2015). Additionally, open-ended protocols used in peer observations may be too general, hindering specific constructive feedback about evidence-based teaching practices. For example, many official peer observations forms contain vague statements such as "Comment on student involvement and interaction with the instructor" or ask whether the instructor was "well prepared" (Table 5; Millis, 1992; Brent and Felder, 2004). When we compare these types of questions with the components of teaching that are best evaluated by peers (Cohen and McKeachie, 1980), we see little overlap (Table 5). This observation has resulted in many universities encouraging the use of collaborative processes rather than simple peer observations (Table 6). Researchers have demonstrated that peers are more likely to reach a stable consensus in reviews of teaching when judgments are based on teaching portfolios (Cohen and McKeachie, 1980; Centra, 2000), which more than 30% of our faculty reported using.

Faculty commonly reported that their peer observers tend to be more positive than their student evaluations. Students and faculty have very different perspectives and goals for teaching (Stark and Freishtat, 2014). In terms of rigor, students have been shown to reward faculty on end-of-semester course evaluations for awarding higher grades (Nimmer and Stone, 1991; Rodabaugh and Kravitz, 1994; Sailor *et al.*, 1997; Carrell and West, 2010). However, two studies that followed up on these same students in subsequent courses found a negative correlation between how they rated faculty and achievement in later courses (Carrell and West, 2010; Braga *et al.*, 2014), implying that students may be negatively rating the most rigorous,

TABLE 5. Revising peer observation forms

Traditional peer observation items (Brent and Felder, 2004)	Providing feedback using components most suitable for peers to assess (Cohen and McKeachie, 1980)
<ul style="list-style-type: none"> • Instructor was well prepared for class. • Instructor was knowledgeable about the subject matter. • Instructor was enthusiastic about the subject matter. • Instructor spoke clearly, audibly, and confidently. • Instructor used a variety of relevant illustrations/examples. • Instructor made effective use of the board and/or visual aids. • Instructor asked stimulating and challenging questions. • Instructor effectively held the class's attention. • Instructor achieved active student involvement. • Instructor treated students with respect. • What worked well in class? • What could have been improved? 	<ul style="list-style-type: none"> • Comment on the appropriateness of the content presented and the learning objectives for students taking this course. • Describe how the instructional materials and organization supported the attainment of these objectives. • Which of the methods used were most successful at achieving the objectives? • Looking over the student performance data included in this evaluation, did the assessments adequately measure these objectives? Were they appropriate for students taking this course? • How did you perceive concern for student learning?

demanding, but yet effective instructors. Alternatively, it is possible that faculty do not provide critical feedback for fear of negative repercussions to the faculty being observed. This implies an implicit decision to give positive, noncritical peer observations as they are often tied to tenure and promotion. While our data clearly show that, for the most part, faculty believe that peer observations are real opportunities for feedback, rather than rubber stamps, it is possible that faculty do not "dig deep" in their peer observations. This means that faculty may not be aware of problems in their classrooms and may be further discounting students' valid comments. Moreover, faculty indicate that what is currently missing from feedback is more constructive feedback. More concrete and specific peer observations using resources from biology education research (e.g., teaching practices inventories) could be used to give faculty the tools to support student buy-in of evidence-based teaching approaches and to be successful in implementing these teaching models (Wieman and Gilbert, 2014; Eddy *et al.*, 2015). Ideally, peer observers should work with faculty to reconcile discrepancies in feedback from classroom observations and student evaluations. Understanding what faculty identified as missing is critical as we move forward to develop new systems for formative feedback about teaching. Several groups have proposed better models for just this type of formative peer evaluation (Table 6; Franchini, 2008; Pressick-Kilborn and te Riele, 2008).

We also need to consider whether it is likely that instructional feedback alone could drive lasting changes in faculty teaching practices. Bouwma-Gearhart (2012) reports that faculty motivation to change their teaching practices may be firmly rooted in extrinsic motivation to preserve their professional

TABLE 6. Peer evaluation resources

The Department Chair's Role in Developing New Faculty into Teachers and Scholars (Bensimon *et al.*, 2000) provides suggestions for assigning teaching mentors to new faculty and methods to provide formative feedback and guidance, including providing new faculty with information that assists them in planning and adapting courses after presentation of feedback.

“Survey of 12 Strategies to Measure Teaching Effectiveness” (Berk, 2005) describes and then critically reviews the dozen methods currently available for evaluating teaching, including peer review and student evaluations, but also teaching awards, portfolios, and student performance data.

Assessing and Improving Your Teaching (Blumberg, 2014) outlines a self-reflective approach that focuses on collecting a variety of sources to document teaching effectiveness and contains multiple rubrics to assess how this documented process changes over time.

Peer Review of Teaching: A Sourcebook, Second Edition (Chism, 2007) provides a very practical guide that provides clear descriptions of systems and materials for observing faculty in the classroom and providing feedback as colleagues.

“Peer Observation and Assessment of Teaching” (originally developed and edited in 2006 by Bill Roberson, PhD, for the University of Texas at El Paso; revised and adapted by Franchini, 2008) is a comprehensive manual that covers everything from rationale for peer observation to issues with selecting good peer observers. It also provides documents and indices to assist in selecting a peer review process and then delivering feedback to faculty.

“Collaborative Peer-Supported Review of Teaching” (Gosling, 2014) reviews the different models of peer review that faculty could adopt to inform their teaching, including evaluative, developmental, and collaborative versions.

University websites also contain detailed descriptions of appropriate ground rules for conducting peer review of class teaching that involve much more than simple drop-in observations and creation of reports:

- University of Arizona (Novodvorsky, 2016)
- Kansas State University (Northway, 2015)
- Vanderbilt University (Bandy, 2015)
- University of Nebraska–Lincoln (2016)

ego, yet these extrinsic motivations shift to more intrinsic motivations over time. Our findings support this model. In fact, when asked what would motivate them to improve their teaching (note: this may or may not be related to implementing evidence-based teaching practices), faculty often cited extrinsic incentives including money, promotion, and awards. Yet when asked what realized rewards they accrued from improving their teaching, faculty more frequently reported intrinsic rewards such as higher-quality student assignments and improved student rapport, with less focus on the extrinsic rewards. Like Bouwma-Gearhart’s findings, our data indicate that the benefits faculty actually accrue are acutely different from the incentives that might initially motivate them to make changes. Future research may explore whether higher education transformation should emphasize a cultural change to increase the extrinsic value placed on teaching quality, or whether the promotion of intrinsic rewards alone would effectively convince faculty to implement evidence-based teaching practices.

We learned that faculty seek additional formative feedback from many sources, including the education research litera-

ture and teaching workshops, to bridge the instructional-feedback gap. By and large, faculty seek out additional feedback from colleagues at their own or other institutions. Based on what we learned from faculty efforts, what can we leverage as new tools to use to work with faculty (to offer formative feedback) rather than against them (evaluative feedback)? From prior work, features of peer observations that encourage faculty to change their teaching practices include: voluntary, formative feedback that is timely, ongoing, self-referenced, and does not threaten the recipient’s self-esteem (Gormally *et al.*, 2014). Additionally, encouraging, positive feedback that accounts for the individual’s confidence and experience leads to higher goal setting. Our current results indicate that faculty may value and use formative feedback received early in the semester rather than summative feedback received after the course has ended. Other work supports this approach. For example, faculty were more likely to decide to alter their teaching practices during or immediately after teaching a particular class session (65%) as opposed to during the following class session or the next time they taught a course (McAlpine and Weston, 2000). Further, McAlpine and Weston explained that faculty have a “corridor of tolerance,” meaning that many aspects of teaching are not modified as long as the teaching factors of interest fall inside what is deemed to be acceptable, that is, inside the “corridor of tolerance.” Specifically, faculty modify aspects of teaching in response to cues from students (McAlpine and Weston, 2000), essentially evaluating their expectations in comparison with the classroom reality. McAlpine and Weston hypothesize that when faculty identify problems that are outside this corridor of tolerance, they make decisions that lead to adjustments in actions. McAlpine and Weston’s findings may explain why faculty are more likely to make modifications following negative feedback than positive feedback. To better support faculty implementing evidence-based teaching practices, we recommend that faculty leverage specific tools to consistently define and clarify this “corridor of tolerance” in peer observations. We also recommend peer observers consider the following caveats: the faculty member’s experience teaching this course; his or her familiarity with the teaching approach; whether the faculty member is able to assess students’ understanding; and, finally, whether the faculty member has control over other variables influencing his or her teaching (e.g., class size, support, classroom environment). Faculty peer observations must capitalize on providing feedback to support the moment-to-moment decision making in the classroom. Because faculty respondents sought more rather than less feedback, we also suggest exploring peer-coaching models as potential strategies to provide formative feedback (Huston and Weaver, 2008; Gormally *et al.*, 2014).

Faculty value feedback when it is received from a respected feedback provider (Gormally *et al.*, 2014). Academic colleagues have been shown to play a key role in mentoring in general (particularly for providing emotional support for junior faculty) and by providing opportunities for career advancement (Peluchette and Jeanquart, 2000; van Emmerik and Sanders, 2004). Colleagues have already been shown to be a more important source of information about teaching innovations than formal presentations or literature (Borrego *et al.*, 2010). Further, Andrews and Lemons (2015) discovered that

interpersonal relationships, such as with faculty peers, play a critical role in influencing faculty decisions to change teaching practices, as well as student perceptions. Similarly, our findings show that faculty see their peers as a valuable source of information about teaching and currently seek more feedback from peers to bridge the gap in what is missing. Interestingly, our current findings illuminate new roles and respect for science faculty with education specialties (SFES; Bush *et al.*, 2006). Survey data revealed that faculty frequently mentioned they would select faculty with education research expertise as feedback providers and that education research expertise was a characteristic they sought when seeking out further feedback. Biologists have clearly come to value what SFES can offer to their departments (Bush *et al.*, 2013). For SFES, our findings about professional development suggest that efforts should focus more on instructional feedback and less on disseminating evidence-based teaching practices. Although many research groups have begun to characterize some of the potential roles for faculty in mentoring peers to improve undergraduate teaching, no one has investigated how faculty interact with colleagues in a mentoring capacity to improve undergraduate teaching. Alternatively, many universities have teaching and learning centers with professionals who have education research expertise. Because these individuals are situated outside academic departments, they are in a position to offer individualized, constructive, nonevaluative, and thus ego-saving, feedback.

Our research reveals faculty-vetted ideas that could be leveraged for new models of instructional feedback. Our data show that faculty have already developed their own “work-around” strategies to address unmet needs for better instructional feedback. Importantly, faculty still highly value feedback from students but recommend using student assessment data and data from student learning in subsequent courses to more effectively improve teaching, rather than end-of-course evaluations alone. For professional development, this may mean a shift from disseminating the gospel of evidence-based teaching practices to a new focus on supporting faculty through the life cycle of implementing and improving these practices. Professional development efforts emphasizing constructive instructional feedback as a “lifestyle” for undergraduate STEM education may prove more effective than isolated training experiences or onetime workshops.

ACKNOWLEDGMENTS

We acknowledge continuing support and feedback from the University of Georgia SEER (Scientists Engaged in Education Research) Center and faculty who helped pilot test the survey. Evan Conaway provided invaluable Qualtrics advice and support. A.M.M. received funding as an Undergraduate Biology Education Research fellow through National Science Foundation Research Experiences for Undergraduates grant #1262715. Funding for data analysis from the survey was provided by the University of Georgia Science Technology Engineering and Math Initiative Small Grants Program. Permissions to survey biology faculty and receive consent to use the results of the survey were obtained from the University of Georgia Institutional Review Board (STUDY00001138) and Gallaudet University Institutional Review Board (2452).

REFERENCES

- Aleamoni LM (1999). Student Rating Myths versus Research Facts from 1924 to 1998. *J Pers Eval Educ* 13, 153–166.
- American Association for the Advancement of Science (2010). *Vision and Change: A Call to Action*, Washington, DC.
- Anderson K, Miller ED (1997). Gender and student evaluations of teaching. *PS: Political Science and Politics* 30, 216–219.
- Anderson WA, Banerjee U, Drennan CL, Elgin SCR, Epstein IR, Handelsman J, Hatfull GF, Losick R, O’Dowd DK, Olivera BM, *et al.* (2011). Changing the culture of science education at research universities. *Science* 331, 152–153.
- Andrews TC, Lemons PP (2015). It’s personal: biology instructors prioritize personal evidence over empirical evidence in teaching decisions. *CBE Life Sci Educ* 14, ar7.
- Andrews TM, Leonard MJ, Colgrove CA, Kalinowski ST (2011). Active learning NOT associated with student learning in a random sample of college biology courses. *CBE Life Sci Educ* 10, 394–405.
- Association of American Universities (2014). Undergraduate STEM Initiative, Washington, DC. <https://stemedhub.org/groups/aau> (accessed July 2015).
- Atwood CH, Taylor JW, Hutchings PA (2000). Why are chemists and other scientists afraid of the peer review of teaching? *J Chem Educ* 77, 239–243.
- Austin AE (2002). Preparing the professoriate of the future: graduate student socialization for faculty role. In: *Higher Education: Handbook of Theory and Research*, ed. JC Smart, Secaucus, NJ: Springer, 397–456.
- Ballantyne R, Borthwick J, Packer J (2000). Beyond student evaluation of teaching: identifying and addressing academic staff development needs. *Assess Eval High Educ* 25, 221–236.
- Bandy J (2015). *Peer Review of Teaching*, Nashville, TN: Vanderbilt University Center for Teaching and Learning.
- Basow SA (1995). Student evaluations of college professors: when gender matters. *J Educ Psychol* 87, 656–665.
- Bell A, Mladenovic R (2008). The benefits of peer observation of teaching for tutor development. *High Educ* 55, 735–752.
- Bensimon EM, Ward K, Sanders K (2000). *The Department Chair’s Role in Developing New Faculty into Teachers and Scholars*, Bolton, MA: Anker.
- Berk RA (2005). Survey of 12 strategies to measure teaching effectiveness. *Int J Teach Learn High Educ* 17, 48–62.
- Berstein DJ (2008). Peer review and evaluation of the intellectual work of teaching. *Change* 40, 48–51.
- Blumberg P (2014). *Assessing and Improving Your Teaching*, San Francisco, CA: Jossey-Bass.
- Boring A, Ottoboni K, Stark PB (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*, DOI: 10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1.
- Borrego M, Froyd JE, Hall TS (2010). Diffusion of engineering education innovations: a survey of awareness and adoption rates in U.S. engineering departments. *J Eng Educ* 99, 185–207.
- Bouwma-Gearhart J (2012). Research university STEM faculty members’ motivation to engage in teaching professional development: building the choir through an appeal to extrinsic motivation and ego. *J Sci Educ Tech* 21, 558–570.
- Bradford SE, Miller ER (2015). Improve undergraduate science education. *Nature* 523, 282–284.
- Braga M, Paccagnella M, Pellizzari M (2014). Evaluating students’ evaluations of professors. *Econ Educ Rev* 41, 71–88.
- Brent R, Felder RM (2004). A Protocol for Peer Review of Teaching. Paper presented at American Society for Engineering Education Annual Conference and Exposition; September 20–October 1, 2004, Kansas Technology Center.
- Brickman P, Gormally C, Armstrong N, Hallar B (2009). Effects of inquiry-based learning on students’ science literacy skills and confidence. *Int J Scholarsh Teach Learn* 3(2).
- Bush SD, Pelaez NJ, Rudd JA, Stevens MT, Tanner KD, Williams KS (2013). Widespread distribution and unexpected variation among science faculty with education specialties (SFES) across the United States. *Proc Natl Acad Sci USA* 110, 7170–7175.
- Bush SD, Pelaez NJ, Rudd JA, Stevens MT, Williams KS, Allen DE, Tanner KD (2006). On hiring science faculty with education specialties for your science (not education) department. *Cell Biol Educ* 5, 297–305.

- Carrell SE, West JE (2010). Does professor quality matter? Evidence from random assignment of students to professors. *J Polit Econ* 118, 409–432.
- Cashin WE (1990). Students do rate academic fields differently. In: *New Directions for Teaching and Learning*, ed. M Theall and J Franklin, San Francisco, CA: Jossey-Bass, 113–121.
- Centra JA (1993). *Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness*, Jossey-Bass Higher and Adult Education Series, San Francisco, CA: Jossey-Bass, 245.
- Centra JA (2000). Evaluating the teaching portfolio: a role for colleagues. *New Dir Teach Learn* 2000, 87–93.
- Chase A, Pakhira D, Stains M (2013). Implementing process-oriented, guided-inquiry learning for the first time: adaptations and short-term impacts on students' attitude and performance. *J Chem Educ* 90, 409–416.
- Chism NV (2007). *Peer Review of Teaching: A Sourcebook*, 2nd ed., Bolton, MA: Anker.
- Cohen PA, McKeachie WJ (1980). The role of colleagues in the evaluation of college teaching. *Improv Coll Univ Teach* 28, 147–154.
- Collins D (2003). Pretesting survey instruments: an overview of cognitive methods. *Qual Life Res* 12, 229–238.
- Cosh J (1998). Peer observation in higher education—a reflective approach. *Innov Educ Train Int* 35, 171–176.
- Cramer KM, Alexitch LR (2000). Student evaluations of college professors: identifying sources of bias. *Can J Hig Educ* 30, 143–164.
- Ebert-May D, Derting TL, Henkel TP, Maher JM, Momsen JL, Arnold B, Passmore HA (2015). Breaking the cycle: future faculty begin teaching with learner-centered strategies after professional development. *CBE Life Sci Educ* 14, ar22.
- Eddy SL, Converse M, Wenderoth MP (2015). PORTAAL: a classroom observation tool assessing evidence-based teaching practices for active learning in large science, technology, engineering, and mathematics classes. *CBE Life Sci Educ* 14, ar23.
- Ericsson KA, Simon HA (1980). Verbal reports as data. *Psychol Rev* 87, 215–251.
- Fairweather J (2009). *Linking Evidence and Promising Practices in Science, Technology, Engineering, and Mathematics (STEM) Undergraduate Education*, Washington, DC: National Academies National Research Council Board of Science Education, 31.
- Feldman KA (1993). College students' views of male and female college teachers: part II—evidence from students' evaluations of their classroom teachers. *Res High Educ* 34, 151–211.
- Foote KE (2010). Creating a community of support for graduate students and early career academics. *J Geogr High Educ* 34, 7–19.
- Franchini B (2008). *Peer Observation and Assessment of Teaching* (adapted with permission from the University of Texas El Paso), New York: University of Albany.
- Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP (2014). Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci USA* 111, 8410–8415.
- Freiberg HJ (1987). Enriching feedback to student-teachers through small group discussion. *Teach Educ Q* 14, 71–82.
- Gormally C, Sullivan CS, Szeinbaum N (2016). Uncovering barriers to enacting inquiry teaching: inconsistent facilitation techniques, student resistance, and reluctance to share control over learning with students. *J Microbiol Biol Educ* 17, 215–224.
- Gormally C, Evans M, Brickman P (2014). Feedback about teaching in higher ed: neglected opportunities to promote change. *CBE Life Sci Educ* 13, 187–199.
- Gosling D (2014). Collaborative peer-supported review of teaching. In: *Peer Review of Learning and Teaching in Higher Education: International Perspectives*, ed. J Sachs and M Parsell, Houten, Netherlands: Springer, 13–31.
- Henderson C, Dancy M (2008). Physics faculty and educational researchers: divergent expectations as barriers to the diffusion of innovations. *Am J Phys* 76, 79–91.
- Henderson C, Dancy MH (2009). Impact of physics education research on the teaching of introductory quantitative physics in the United States. *Phys Rev Spec Top Phys Educ Res* 5, 020107.
- Henderson C, Dancy M, Niewiadomska-Bugaj M (2012). Use of research-based instructional strategies in introductory physics: where do faculty leave the innovation-decision process? *Phys Rev Spec Top Phys Educ Res* 8, 020104.
- Henderson C, Turpen C, Dancy M, Chapman T (2014). Assessment of teaching effectiveness: lack of alignment between instructors, institutions, and research recommendations. *Phys Rev Spec Top Phys Educ Res* 10, 010106.
- Hoffmann F, Oreopoulos P (2009). Professor qualities and student achievement. *Rev Econ Stat* 91, 83–92.
- Huston T, Weaver CL (2008). Peer coaching: professional development for experienced faculty. *Innov High Educ* 33, 5–20.
- Keig L (2000). Formative peer review of teaching: attitudes of faculty at liberal arts colleges towards colleague assessment. *J Pers Eval Educ* 14, 67–87.
- Kell C, Annetts S (2009). Peer review of teaching embedded practice or policy-holding complacency? *Innov Educ Teach Int* 46, 61–70.
- Kember D, Leung DYP, Kwan KP (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assess Eval High Educ* 27, 411–425.
- Loeher L (2006). *An Examination of Research University Faculty Evaluation Policies and Practice*, Portland, OR: Professional and Organizational Development.
- Marsh HW, Roche L (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *Am Educ Res J* 30, 217–251.
- Martin GA, Double JM (1998). Developing higher education teaching skills through peer observation and collaborative reflection. *Innov Educ Train Int* 35, 161–170.
- McAlpine L, Weston C (2000). Reflection: issues related to improving professors' teaching and students' learning. *Instr Sci* 28, 363–385.
- McCormick AC (2015). *The Carnegie Classification of Institutions of Higher Education*, Bloomington: University of Indiana.
- Millis BJ (1992). Conducting effective peer observations. *To Improve the Academy* 11, 189–206.
- Murray HG (1983). Low-inference classroom teaching behaviors and student ratings of college teaching effectiveness. *J Educ Psychol* 75, 138–149.
- National Center for Education Statistics (2011). *Digest of Education Statistics*, Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.
- National Research Council (2015). *Reaching Students: What Research Says About Effective Instruction in Undergraduate Science and Engineering*, Washington, DC: National Academies Press.
- Nielsen HD, Moos RH, Lee EA (1978). Response bias in follow-up studies of college students. *Res High Educ* 9, 97–113.
- Nimmer JG, Stone EF (1991). Effects of grading practices and time of rating on student ratings of faculty performance and student learning. *Res High Educ* 32, 195–215.
- Northway K (2015). *Peer Review of Teaching Program*, Web page. Kansas State University.
- Novodvorsky I (2016). *Peer Review of Teaching Protocol*, Tucson, AZ: Office of Instruction and Assessment in the Manuel Pacheco Integrated Learning Center.
- Peel D (2005). Peer observation as a transformatory tool? *Teach High Educ* 10, 489–504.
- Peluchette JV, Jeanquart S (2000). Professionals' use of different mentor sources at various career stages: implications for career success. *J Soc Psychol* 140, 549–564.
- Pershing JA (2006). *Handbook of Human Performance Technology: Principles, Practices, and Potential*, 3rd ed., San Francisco, CA: Pfeiffer.
- President's Council of Advisors on Science and Technology (2012). *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*, Washington, DC: U.S. Government Office of Science and Technology, 130.
- Pressick-Kilborn K, te Riele K (2008). Learning from reciprocal peer observation: a collaborative self-study. *Studying Teach Educ* 4, 61–75.
- Qualtrics (2016). www.qualtrics.com.
- Quinlan K, Bernstein DJ (1996). Special Issue on peer review of teaching. *Innov High Educ* 20, 219–307.

- Richardson JTE (2005). Instruments for obtaining student feedback: a review of the literature. *Assess Eval High Educ* 30, 387–415.
- Richlin L, Essington A (2004). Faculty learning communities for preparing future faculty. *New Dir Teach Learn* 2004, 149–157.
- Rodabaugh RC, Kravitz DA (1994). Effects of procedural fairness on student judgments of professors. *J Excell Coll Teach* 5, 67–83.
- Sailor P, Worthen BR, Shin E-H (1997). Class level as a possible mediator of the relationship between grades and student ratings of teaching. *Assess Eval High Educ* 22, 261–269.
- Seldin P (1999). *Changing Practices in Evaluating Teaching: A Practical Guide to Improved Faculty Performance and Promotion/Tenure Decisions*, Boston, MA: Anker.
- Shortland S (2010). Feedback within peer observation: continuing professional development and unexpected consequences. *Innov Educ Teach Int* 47, 295–304.
- Smith MK, Jones FHM, Gilbert SL, Wieman CE (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. *CBE Life Sci Educ* 12, 618–627.
- Stark PB, Freishtat R (2014). An evaluation of course evaluations. *Science-Open*, DOI: 10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1.
- Stes A, Min-Leliveld M, Gijbels D, Van Petegem P (2010). The impact of instructional development in higher education: the state-of-the-art of the research. *Educ Res Rev* 5, 25–49.
- Thomas S, Chie QT, Abraham M, Raj SJ, Beh LS (2014). A qualitative review of literature on peer review of teaching in higher education: an application of the SWOT framework. *Rev Educ Res* 84, 112–159.
- Turpen C, Finkelstein ND (2009). Not all interactive engagement is the same: variations in physics professors' implementation of "peer instruction." *Phys Rev Spec Top Phys Educ Res* 5, 020101.
- Turpen C, Finkelstein ND (2010). The Construction of Different Classroom Norms during Peer Instruction: Students Perceive Differences. *Phys Rev Spec Top Phys Educ Res* 6, 020123.
- University of Nebraska–Lincoln (2016). Peer Review of Teaching Project. <http://peerreview.unl.edu> (accessed 16 November 2016).
- van Emmerik H, Sanders K (2004). Social embeddedness and job performance of tenured and non-tenured professionals. *Hum Resour Manage J* 44, 40–54.
- Wachtel HK (1998). Student evaluation of college teaching effectiveness: a brief review. *Assess Eval High Educ* 23, 191–211.
- Walker JD, Cotner SH, Baepler PM, Decker MD (2008). A delicate balance: integrating active learning into a large lecture course. *CBE Life Sci Educ* 7, 361–367.
- Watkins D, Hattie J (1985). A longitudinal-study of the approaches to learning of Australian tertiary students. *Hum Learn* 4, 127–141.
- Weimer M (2002). *Learner-Centered Teaching*, San Francisco, CA: Jossey-Bass.
- White J, Pinnegar S, Esplin P (2010). When learning and change collide: examining student claims to have "learned nothing." *J Gen Educ* 59, 124–140.
- Wieman C, Gilbert S (2014). The Teaching Practices Inventory: a new tool for characterizing college and university teaching in mathematics and science. *CBE Life Sci Educ* 13, 552–569.