

# Faculty Perspectives about Instructor and Institutional Assessments of Teaching Effectiveness

Chandra Turpen\*, Charles Henderson\*, and Melissa Dancy<sup>†</sup>

*\*Department of Physics and Mallinson Institute for Science Education, Western Michigan University, Kalamazoo, MI 49008, USA*

*<sup>†</sup>Department of Physics, University of Colorado, Boulder, CO 80309, USA*

**Abstract.** Faculty and their institutions should have a shared set of metrics by which they measure teaching effectiveness. Unfortunately, the current situation at most institutions is far from this ideal. As part of a larger interview study, physics faculty were asked to describe how they and their institutions evaluate teaching effectiveness. Institutions typically base most or all of their assessment of teaching effectiveness on the numerical ratings from student evaluations of teaching effectiveness. Faculty, on the other hand, base most or all of their assessment of teaching effectiveness on student test performance and ongoing formative assessments. In general, faculty are much more positive about the methods that they use to evaluate their teaching than the methods that their institution uses to evaluate their teaching.

**Keywords:** assessment of teaching effectiveness, educational change

**PACS:** 01.40.Fk

## INTRODUCTION

Strong assessment methods are broadly recognized as integral to effective instruction [1]. Three current trends in higher education are encouraging institutions to pay more attention to assessing student learning outcomes: increasing accountability pressures, increased competition for students, and increased use of evidence-based management techniques [Ref. 2, pg. 7]. Thus, since there is an interest in changing assessment practices, now is an ideal time to influence practice. The purpose of this paper is to present the perspectives that faculty have about the methods that both they and their institutions use to evaluate teaching effectiveness.

It is difficult to talk about measures of teaching effectiveness without talking about Student Evaluations of Teaching (SETs). SETs are the most common instrument used in the institutional assessment of teaching [3, 4]. There is a large body of literature and considerable disagreement about the value of SETs. The purpose of this paper is not to enter this debate. Thus, we will simply acknowledge that some researchers argue that SETs are valid since SETs are correlated with student achievement [4, 5]. Other researchers argue that the SETs are merely popularity contests and that the use of SETs is a barrier to more effective teaching, since faculty are

hesitant to be more rigorous due to fears of lower SETs [3, 6, 7]. There is almost universal agreement, however, that assessment of something as complicated, nuanced, and important as teaching effectiveness should be made using multiple methods [e.g., 3, 5, 8, 9]. We suspect that SETs are an appropriate part of a complete assessment system.

With respect to educational change, authors have argued that changing assessment methods will be instrumental in promoting more innovative, research-proven teaching practices [9]. In addition to serving as potential levers for instigating change, the assessment methods used by institutions and instructors to determine teaching effectiveness will influence instructors' determinations about whether their attempts to use research-based instructional strategies (RBIS) are working. In this way, assessment practices influence judgments about the relative advantage and compatibility [10] of educational innovations and thus the continuation or discontinuation of RBIS use.

## METHODS

In the Fall of 2008, a sample of physics faculty from across the country were asked to complete a survey about their instructional goals and practices as well as their knowledge and use of RBIS (see Ref. 11 for more details). Respondents included faculty from

**TABLE 1.** Descriptions of each source of assessment information.

Sources of Assessment Information
<b>Student Evaluations of Teaching:</b> All structured collection of student evaluative feedback about a course.
<b>Peer Observations of Teaching:</b> Having peer faculty or university administrators sit in on an instructor's course and provide feedback (often through written reports and sometimes verbal feedback in a face-to-face meeting).
<b>Teaching Portfolios:</b> Having instructors self-report how they teach, sometimes providing references to the research literature about the evidence for the success of the instructional methods they are using.
<b>Research-based assessments:</b> Typically involving pre/post-testing through the use of multiple-choice conceptual inventories.
<b>Student Performance on Exams, Quizzes, or Homework:</b> Using students' performance on exams, quizzes, or homework and taking this performance as an indicator of the success of instruction.
<b>Systematic Formative Assessment:</b> Gathering a sampling of students' performance on an in-class task as an indicator of teaching effectiveness (e.g., having students submit votes, walking around the room and observing the solutions of multiple groups).
<b>Informal Formative Assessment:</b> All other forms of formative assessment, such as students' verbal comments in class or office hours, the look of confusion in students' eyes, whether or not students are awake, or whether or not students are asking questions.
<b>Informal Post-course Feedback from Students:</b> Spontaneous informal comments from student(s) semesters or years after the student(s) have left the class with feedback on the course or the instructor.

both four- and two-year institutions. The overall response rate for the survey study was 50.3%. A subset of survey respondents was purposefully chosen to participate in an associated interview study. Of the 51 faculty contacted for interviews about 69% (N=35) agreed to participate in the interview study.

Semi-structured interviews were conducted with 35 physics faculty with a variety of experiences with Peer Instruction [12]. Across institution types, we interviewed faculty from each of three self-reported user categories: User, Former User, and Knowledgeable Non-user. Although the interviews were primarily focused on knowledge about and use of Peer Instruction, most participants were asked to explicitly discuss assessment issues through the following interview questions: A) How do you know if your instruction is working? B) What evaluation criteria does your institution use to evaluate teaching? C) Is there an observational component to your teaching evaluation? If so, are there any specific criteria or specific behaviors that are being looked for in these observations? and D) Do you receive feedback on your teaching from your students? What kinds of things do students use to evaluate good teaching? Interviews typically lasted over one hour and were audio-recorded and transcribed.

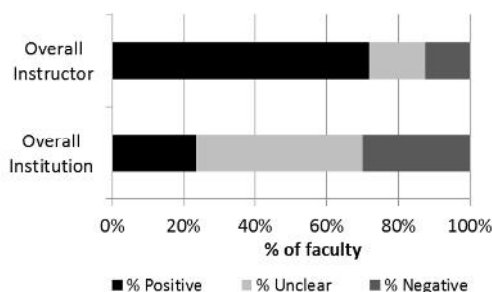
For the purposes of understanding interviewees' views about assessment of teaching, each interview was first coded using a broad category of "commentary on assessment." This code captured all classroom scale and institutional level assessment practices, i.e. reading the expressions on students' faces in class to determine instructional moves, using students' answers to in-class questions to determine how to proceed, or the institutional use of SETs to determine the teaching component of promotion and tenure. The quotations in this broad category were iteratively revisited to develop a more specific coding scheme consisting of eight common sources of assessment information (Table 1). With respect to each

source, during the analysis we sought to answer the following questions: Was this source mentioned as possibly relevant for an instructor's (or institution's) assessment of teaching? Was this source of information used by the instructor (or institution) to assess teaching effectiveness? Did the interviewee see value in their own use (or the institution's use) of this source of information? Additionally, the researchers holistically assessed whether the interviewee felt positively, negatively or ambivalent about how well they are able to evaluate their own teaching effectiveness. Similarly, the researchers holistically assessed whether the interviewee felt positively, negatively, or ambivalent about how well they think their institutions are able to evaluate teaching effectiveness. Because these driving analytical questions arose from the analysis of the interviews and were not asked directly in the interview, there were many instances in which there was insufficient data upon which to answer for particular faculty.

## DATA & RESULTS

We begin by examining interviewee's holistic impressions about the success of instructor and institutional assessment practices (as shown in Figure 1). Of the 32 faculty who discussed their assessment of teaching, most (~72%) were satisfied with their efforts to assess whether their instruction is working. However, of the 30 faculty that discussed institutional assessment practices, only a small minority of faculty (~23%) were satisfied with their institution's assessment of teaching effectiveness. A significant fraction of instructors (~30%) were explicitly negative about how their institution assesses teaching effectiveness.

In this way, faculty seem to have more confidence in their own assessment practices than they do in the assessment practices of the institution. For example,



**FIGURE 1.** Interviewees' sense of how well they and their institutions are able to assess teaching effectiveness.

one interviewee said, "it's the professor's job to figure out how that they can best learn. But until you're tenured the reality is, you know, your evaluations had best sing" (PUBT2).

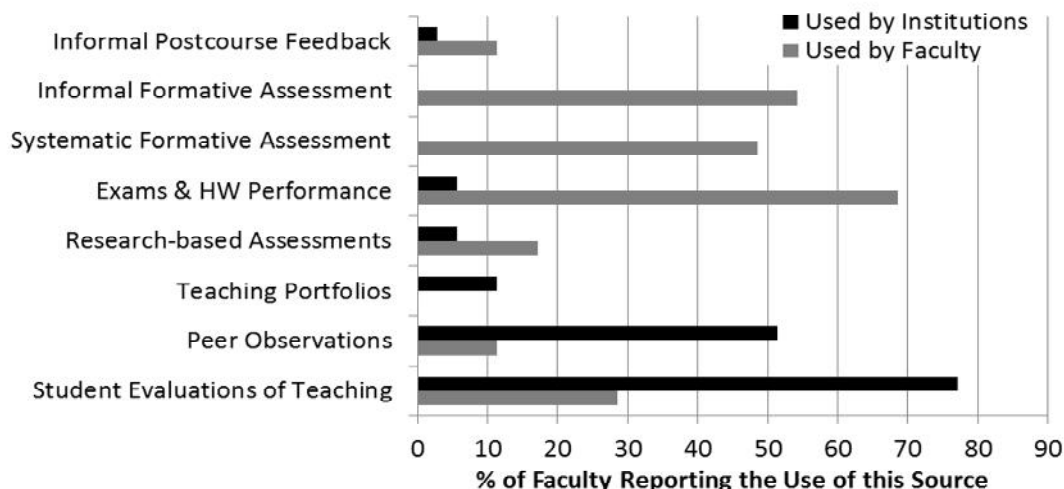
We find very little overlap in the prevalent sources of assessment information used by faculty and the prevalent sources of assessment information reportedly used by institutions (as shown in Figure 2). Faculty report that institutions tend to use SETs (77%) and peer observations of teaching (51%) to judge teaching effectiveness. Faculty most commonly use 1) students' performance on exams, quizzes, and homework (69%), 2) informal formative assessment (54%), and 3) systematic formative assessment (49%) to judge teaching effectiveness.

Even the apparent overlap in use of SETs by institutions and faculty is somewhat misleading. Faculty often report administering their own, more informal SETs in parallel with the university's efforts, but with different questions for their students. Even the faculty that do perceive some value in the university-administered SETs typically skim (or

ignore) the multiple-choice questions on the survey and mostly (or entirely) focus on students' long answer comments.

Thirty of the 35 faculty interviewed explicitly discussed some aspects of institutional assessment practices. Of these 30 who discussed institutional assessment, one third described the institution as only using one measure of teaching effectiveness. SETs were reported to be the most common single source that institutions relied upon (8/10). An additional 15 faculty described their institutions as only relying on two assessment measures for judging teaching effectiveness. The most common combination of two assessment measures for institutions was SETs paired with peer observations (12/15). Our preliminary findings suggest that even institutions that rely on two measures tend to give significant (or sole) prominence to SETs.

Our interview sample was specifically chosen based on some familiarity with Peer Instruction. However, even within this population we find a relatively low use of research-based assessments to judge teaching effectiveness (by either individual instructors or institutions). It is unclear the extent to which this might be a principled or reasoned decision to not use these instruments or if there is a lack of awareness about the availability of these instruments. Our preliminary findings suggest that a subset of faculty do not feel that the currently available conceptual inventories measure the range of goals that they have in their introductory courses (such as proficiency at quantitative problem solving, understanding the culture of physics, or critical thinking).



**FIGURE 2.** Reported use of various sources of assessment information by instructor and by institutions in judging teaching effectiveness.

## CONCLUSIONS & IMPLICATIONS

The results presented above can be summarized in four important conclusions related to the assessment of teaching effectiveness:

1. Faculty are much more positive about the methods that they use to evaluate their teaching than the methods that their institution uses to evaluate their teaching.
2. Institutions typically base most or all of their assessment of teaching effectiveness on the numerical ratings from SETs, a measure that many faculty are skeptical of.
3. Faculty tend to base most or all of their assessment of teaching effectiveness on student test performance and ongoing formative assessments.
4. Neither faculty nor their institutions make much use of the available nationally-normed research-based assessments (such as the FCI).

There is a drastic discontinuity between the information instructors and institutions are using to judge teaching effectiveness. Both faculty and institutions could benefit from broadening the assessment sources that they use to determine teaching effectiveness—specifically through the use of nationally-normed assessments that would allow for inter-institutional comparison. Instructors may benefit from breaking out of the “private” or isolated nature of their typical classroom assessment methods and engaging in broader conversations about assessment. Through exploring complementary, more standardized measures of teaching effectiveness based on student learning, instructors would be able to compare across classrooms which may offer new insights into the range of achievement that is possible as well as provide insights into instructional strategies or practices worth experimenting with.

Toward this end, we now briefly turn our attention to some speculations about what the educational research community (especially PER) can do to improve this situation. Given the current national emphasis on accountability in higher education, we suggest that this is an ideal time for PER to join the conversation. The PER community can provide a moderating voice that encourages the coordinated use of formative assessment and standardized research-based summative assessments. PER has developed a number of research tools and techniques that are commonly used to assess student learning, but so far PER has not significantly promoted the use of these tools in the assessment of teaching effectiveness. We believe that the PER community is well-situated to a) provide scaffolds that will build on instructors’ current assessment practices based on student performance on

exams and homework, and formative assessment (e.g., research-based rubrics to evaluate student work [13]), and b) encourage the complementary use of standardized assessment measures (such as conceptual inventories or other research-based assessments).

In encouraging the coordination of multiple sources of assessment information, the PER community could also provide guidance in how to integrate and make sense of multiple data sources, particularly in situations in which the data sources provide contradictory information.

## ACKNOWLEDGMENTS

This paper is based upon work supported by the National Science Foundation under Grant No. 0715698. We thank all of the professors who shared their time and their experiences with us.

## REFERENCES

1. J. M. Atkins and J. E. Coffey (eds.), *Everyday Assessment in the Science Classroom*, Arlington, VA: National Science Teachers Association Press, 2003.
2. P. T. Ewell, *Making the Grade: How Boards Can Ensure Academic Quality*, Washington, DC: Association of Governing Boards of Universities, 2006.
3. S. E. Carrell and J. E. West, “Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors.” *Journal of Political Economy*, **118**(3), 409-432 (2010).
4. H. Wachtel, “Student Evaluation of College Teaching Effectiveness: a brief review.” *Assessment & Evaluation in Higher Education*, **23**(2), 191-212 (1998).
5. K. A. Feldman, “Identifying Exemplary Teachers and Teaching: Evidence from Student Ratings. In R. P. Perry & J. C. Smart (Eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-based Perspective* (pp. 93-143). Dordrecht: Springer (2007).
6. R. Arum and J. Roska, *Academically Adrift*. Chicago: University of Chicago Press, 2011.
7. V. E. Johnson, *Grade Inflation: A Crisis in College Education*. New York: Springer (2003).
8. W. A. Anderson et al., *Science* **331**(6014), 152-153 (2011).
9. E. Seymour, “Tracking the processes of change in US undergraduate education in science, mathematics, engineering, and technology,” *Science Education*, **86**(1), 79-105 (2002).
10. E. M. Rogers, *Diffusion of Innovations*, 5<sup>th</sup> ed., New York: Free Press, 2003.
11. C. Henderson and M. Dancy, *Phys. Rev. Phys. Rev. ST Physics Ed. Research* **5**, 020107 (2009).
12. E. Mazur, *Peer Instruction: A User’s Manual*, Upper Saddle River NJ: Prentice Hall, 1997.
13. E. Etkina et al., *Phys. Rev. Phys. Rev. ST Physics Ed. Research* **5**, 010109 (2009).