

**Perceptions of Administrators, Associate Instructors, Faculty, and Students about
the Evaluation of Teaching in the School of Education: A Report for the
Committee on Teaching**

Executive Summary

The Committee to Review the Evaluation of Teaching, a sub-committee of the Committee on Teaching, was composed of a group of faculty and associate instructors who came together with common interests and questions about the way teaching is evaluated in the School of Education. Clearly, the promotion of strong teaching practices that foster student learning aligns with the mission of the School of Education. Teaching evaluations are conducted for multiple purposes and the results may be employed in different ways. We viewed a review of practices for the evaluation of teaching in the School of Education as timely.

Several questions guided our inquiry. These questions related to the ways that various stakeholders with an interest in the evaluation of teaching perceived current practices. To address these questions, we generated a set of survey instruments to elicit data from administrators (i.e., departmental chairpersons), associate instructors, faculty, and students (i.e., members of the Dean=s Advisory Council). The body of this report presents our analysis of the findings.

Based on these findings, we propose the following recommendations for consideration by the Committee on Teaching. Subsequent to their approval of the report, we request that our recommendations be forwarded to Policy Council for its consideration and action.

Recommendations

- Articulate the distinction between evaluating teaching for promotion and tenure and for annual merit review, with respect to the purpose of evaluation as well as the criteria by which an evaluative judgment will be made. The two processes should be made more transparent. Our data also raises the question whether multiple forms of evidence (a more time-intensive evaluative process) should be required for annual merit review as they are now for promotion and tenure.
- Establish common criteria across departments within the School of Education regarding materials required for submission as evidence of quality teaching *and* their weighting.

- Group courses in categories when using student ratings as evaluative data: for example, compare introductory courses with other introductory courses, graduate seminars with other graduate seminars, online courses with online courses, and so forth.
- Delineate the weight given to various forms of documentation.
- Assign less weight to 1st year evaluations.
- Design and implement a coherent and consistent program of guidance and mentoring for associate instructors that is ongoing throughout the academic year. Given the amount of concern about this issue voiced in our data, it seems that the current beginning- of- the- year orientation provided by the Office of Teacher Education is not sufficient. Clearly, associate instructors are requesting support and opportunities to improve their teaching skills.
- Inform associate instructors as to the purpose of evaluating their teaching, the method of doing so, and how these data will be utilized for the purpose of an evaluative decision.
- Clarify to all course instructors the role they can play in constructing the course evaluation instrument (i.e., that they have a choice in selecting which items appear on the BEST evaluation form).
- Advise new instructors to more carefully match BEST items with specific course content and instruction.
- If course evaluations are administered in class, allow adequate time for students to provide thoughtful responses to BEST questions, pending school-wide adoption of online administration of BEST.

**Perceptions of Administrators, Associate Instructors, Faculty, and Students about
the Evaluation of Teaching in the School of Education: A Report for the
Committee on Teaching**

Submitted by the Committee to Review the Evaluation of Teaching,
a Subcommittee of the Committee on Teaching

Rajat Chadha

Judith Chafel (Chairperson)

King, Arthur

Manifold, Marjorie

Roman, Tiffany

Sanchez, Lenny

Spear, Lyndsay

Wall-Marencik, Wendy

March 2, 2009

Introduction and Rationale

The Committee to Review the Evaluation of Teaching is composed of a group of faculty and associate instructors who came together with common interests and questions about the way teaching is evaluated in the School of Education. A relatively large number of people (approximately 20) responded to the call for participation. Many of those who responded and who currently comprise the Committee are associate instructors, suggesting the evaluation of teaching may be of particular interest to graduate students.

Clearly, the promotion of strong teaching practices that foster student learning aligns with the mission of the School of Education. Teaching is, however, a complicated and context-bound process. Consequently, the evaluation of teaching is also complex. The *Promotion and Tenure Criteria for the School of Education* (02.30R, p. 4) appears to acknowledge this complexity in encouraging the use of multiple sources of data in the evaluation of teaching:

Teaching is a complex process that encompasses multiple components, *multiple forms of evidence are needed to assess teaching effectiveness comprehensively*. No one form of evidence (e.g., student evaluations of teaching) should carry disproportionately more weight than any other. *Appropriate* teaching materials may include evidence from the *instructor* (e.g., philosophy of teaching, teaching goals, syllabi, instructional materials, reflections on efforts to evaluate and improve teaching, presentations and articles on one's teaching), evidence from *others* (e.g., colleague evaluations of student outcomes, observations by colleagues trained to evaluate teaching, invitations to share one's teaching expertise), and evidence from *students* (e.g., solicited and unsolicited feedback from students, course-related student products, evidence of student achievement, student-selected teaching awards). These categories of evidence may be interrelated.

(taken verbatim from *Promotion and Tenure Criteria for the School of Education*, 02.30R, p. 4. Available online at:
<http://site.educ.indiana.edu/Portals/28/Policy%20Council/2001-2002/02.30R.pdf>)

There is some question, however, about the extent to which these guidelines are (or can be) carried out in practice.

A vast body of literature debates the value, use, and reliability of student evaluations, and speaks to the worth of implementing a comprehensive system of evaluation comprised of different types of assessment that can be used for a range of purposes. Spencer and Flyr (1992) report that the first teacher

rating scales were published as early as 1915. In the 1920s, student evaluations of teaching were introduced at major universities in the United States (Doyle, 1983; Marsh, 1987). Several researchers (d'Apollonia & Abrami, 1997; Greenwald, 1997; Marsh & Roche, 1997; McKeachie, 1997) agree that course evaluation scores are the only practical indicator of teaching effectiveness with well-established validity of use. Although scholars (e.g., Kulik, 2001; Scriven, 1981) recognize several shortcomings of course evaluations, they are considered to be one of the best and most practical indicators of teaching effectiveness (Scriven, 1989).

Cohen (1981) conducted a meta-analysis of 41 independent validity studies. These studies report a relationship between course evaluation item scores and student achievement in 68 multi-section courses. The average Pearson's Product Moment Correlation between overall instructor rating ("The instructor of this course was outstanding") and student achievement was found to be 0.43. The average correlation between overall course rating ("This was an outstanding course") and student achievement was found to be 0.47. Other methodologically sound studies have yielded relatively consistent findings (Abrami, d'Apollonia, & Cohen, 1990; Feldman, 1989; Kulik, 2001; Marsh, 1984).

Teaching evaluations are conducted for multiple purposes, and the results may be employed in different ways. Instructors, for example, may use evaluation outcomes to diagnose and improve their own instruction. Administrators may use them to make decisions related to hiring, merit, and tenure/promotion processes. Quality assurance panels may use them in assessing and accrediting programs as a whole, and at some universities, students may employ them in the selection of courses (Marsh, 2007, p. 320). Arguably, there is tension among the different uses of evaluation, especially between formative assessments that might be employed to help faculty improve their instruction and summative judgments made in the context of tenure and promotion processes and merit review. There is a need to find means of evaluating teaching that maintain motivations to improve, suggest ways of doing so, yield data to support summative judgments, and yet are simultaneously viewed as legitimate and fair.

We view a review of practices for the evaluation of teaching in the School of Education as timely. What counts as ‘good’ teaching and ‘good’ education changes and evolves over time, and as such the evaluation of teaching should be revisited on a regular basis. Additionally, new technologies now presenting themselves make alternative but relatively untested methods for conducting the evaluation of teaching available. The system of teaching evaluations devised by Bloomington Evaluation Services and Testing (B.E.S.T.) is well established and while it is not our intention to suggest fundamental revisions to this university-wide process, we do consider it worthwhile to explore how those who use this process perceive it. Our objective was to tap the experiences of various constituencies with a stake in the evaluation process in order to formulate recommendations for improvement.

Several questions guided our inquiry:

- How do administrators, associate instructors, faculty, and students perceive the evaluation of teaching in the School of Education?
- To what extent do administrators, associate instructors, and faculty agree that official policies and day-to-day practices related to the evaluation of teaching align?
- To what extent are administrators, associate instructors, faculty, and students satisfied with the current way of evaluating teaching?
- What obstacles do administrators, associate instructors, and faculty perceive in revising the evaluation process? What suggestions do they have for overcoming or resolving these obstacles?
- What suggestions do associate instructors and faculty make for improvement?

Methodology

This section of the report describes the methodology that the Committee devised to address the questions we posed to guide our work. We report on our data sources, instruments, and the procedures we employed to collect the data.

Instruments

In meetings of the Committee, we collectively generated a set of survey instruments to elicit data from constituencies in the School of Education having a stake in the evaluation of teaching: namely, administrators (i.e., departmental chairpersons), associate instructors, faculty, and students (i.e., members of the Dean=s Advisory Council). The Appendix contains a copy of each of these survey instruments.

Procedures for Collecting and Analyzing the Data

The table below shows our data sources, instruments, and modes of data collection. Four departmental chairpersons, 25 associate instructors, 21 regular faculty members, and 2 clinical faculty members responded to our electronic call to complete an on-line survey. The number of regular faculty (21) may include some emeriti faculty. Responses of regular faculty were also recorded at regularly scheduled meetings of each of the five departments in the School of Education (namely, Counseling and Educational Psychology, Curriculum and Instruction, Educational Leadership and Policy Studies, Instructional Systems Technology, and Language Education). At these meetings, a faculty member belonging to the department conducted the interview while a member of the Committee to Review the Evaluation of Teaching recorded the data via handwritten or typed notes. Finally, a member of our committee attended a meeting of the Dean=s Advisory Council and administered a survey to the nine students who attended. Additional data came from an email sent by one faculty member to the chairperson of the Committee. Another faculty member requested a “live interview” with another member of the Committee.

Table 1. Data Sources, Instruments, and Modes of Data Collection

| Data Source | Instrument | Mode of Collection |
|---|-----------------------|---------------------------|
| Administrators (Departmental Chairpersons) | Survey (see Appendix) | Electronic |
| Faculty | Survey (see Appendix) | Departmental Meeting |
| Instructors (Associate Instructors and Faculty) | Survey (see Appendix) | Electronic |
| Students (Dean=s Advisory Council) | Survey (see Appendix) | DAC Focus Group Meeting |

Working independently, four members of the Committee read all four data sets. Working in pairs, they then scrutinized two sets of data more closely. One pair worked with the data collected from instructors and students via an electronic survey as well as from student members of the Dean’s Advisory Council. The other pair read the data collected from departmental chairpersons and from faculty present at departmental meetings. Each member of the pair individually read his or her assigned data to identify major themes. Upon completion of this phase of data analysis, each pair met to compare the themes that each had individually derived inductively. They discussed and compared the themes they found until a consensus was reached. Once the themes had been agreed upon, the two members of each pair then individually analyzed their assigned data to ascertain whether a given theme appeared *more prominently* or *less prominently* in the data. The members of each pair then came together to compare these frequency ratings. The Appendix contains a copy of these analyses for all four data sets. Members of the Committee to Review the Evaluation of Teaching met to review and discuss these findings.

Summary and Discussion of the Findings

By and large, more prominent themes identified from the departmental meeting and departmental chairperson responses supported or were supported by themes emerging from the electronic survey data provided by faculty and associate instructors. Therefore, the data collected from these respondents will be

discussed together. Responses of associate instructors, which were gleaned from the electronic surveys, and those of students who participated in the Dean's Advisory Council focus group meeting are considered separately, as the concerns expressed by these stakeholders presented different points of view from those of faculty members and departmental chairpersons. A summary of findings appears at the end of this section.

Consistency with the Statement

Questions put to department chairpersons, faculty, and associate instructors were prefaced by a statement taken from the *Promotion and Tenure Criteria for the School of Education* (02.30R, p. 4; see Appendix 1). The first and second questions appearing on the survey asked: "In your estimation is the way teaching is evaluated in your department consistent with this statement," and "How does it differ?" Nearly half of the regular faculty and one of the two clinical faculty members responding to the electronic survey indicated that they did not think that the way teaching was evaluated in their department was consistent with the statement. Another 19% were vague or gave mixed responses to this question. Those who responded "no" (i.e., that the evaluation of teaching in their department was not consistent with the statement), expressed strong statements, such as "Absolutely not. The student evaluations are used excessively, despite being misleading in and of themselves;" and, "No, I have never been observed, never been given feedback, and never had someone explain to me the efficacy of the different percentages on the bubble survey (BEST)."

Perceptions of inconsistency between the statement and departmental evaluation practices also appeared in the comments of faculty and chairpersons recorded via the departmental meeting and departmental chairperson surveys. Inconsistencies were noticed more in terms of the ways teaching evaluations were considered for annual merit reviews. A variety of documents are permitted and considered in determining the quality of teaching for promotion and tenure cases where evaluation is criterion-based. Thus, the evaluation process was seen as more in line with the statement. On the other hand, merit reviews were understood as norm-based. Departmental chairpersons and faculty members

indicated “[m]ultiple forms of assessment are not necessarily used for annual review.” Furthermore, lack of standardization regarding other documents which are or are not deemed important for inclusion in annual reports, lack of clarity regarding the weight given various forms of documentation, and differences in terms of materials required by various departments, appeared to result in confusion and initiated questions and concerns among faculty.

How Teaching Is Evaluated in Departments

Student ratings were described by departmental meeting participants as the predominant and most “heavily weighted” instrument used for assessing the quality of teaching. One departmental chairperson observed that the documentation required for annual reviews is narrow, “Best scores, teaching publications, awards. The other things listed above can be included for annual reviews, but they are secondary.” All respondents to the electronic survey also reported that student evaluations were the primary source of data employed for the evaluation of teaching. “Students evaluations carry disproportionate weight”, wrote one respondent. Additionally, eighteen respondents (39%) indicated that peer or mentor observations were considered in assessments of excellent teaching, three (6%) syllabi, and three (6%) personal evaluations or statements as documents considered for this purpose, although these latter evaluative instruments appear to have been relative to promotion and tenure cases or portfolios prepared by associate instructors. At a departmental meeting, one faculty member stated, “I include log data in my annual report, though I don’t think that the school takes that into account.” Another commented, “We do not include things like unsolicited feedback or formal external observations. However, if individuals provide them, they are considered. . . . Most of the evidence relies on the initiative of the instructor – not something that is done by the department or school.”

Satisfaction/Dissatisfaction with Departmental Evaluation of Teaching

Departmental chairpersons, faculty, and associate instructors were asked about their satisfaction or dissatisfaction with their department’s approach to evaluating teaching. Less than 40% of the respondents

to the electronic survey communicated that they were satisfied or mostly satisfied with the departmental approach to evaluating teaching. Comments in support of student ratings conveyed that respondents appreciated the efficiency of the instrument inasmuch as it was readily available, provided global ratings, and required no extra work on the part of instructors. A larger number (57%) of respondents on the electronic survey sent to faculty and associate instructors indicated that they were dissatisfied or had mixed feelings about their department's approach to evaluating teaching. Of these respondents (excluding 6 who signified not applicable), 4 (9%) were completely dissatisfied, 7 (18%) were mostly dissatisfied, and 16 (36%) expressed mixed feelings about the approach their department employed for evaluating teaching for merit review. Reasons given for dissatisfaction or ambiguity about these evaluative processes varied. Although "too much emphasis on this one [BEST] assessment" was a somewhat common theme, others identified the following weaknesses of their current systems: (1) "evaluations are ignored, useless, and do not improve teaching skills," (2) "more flexibility is needed," (3) "more formal time needs to be structured for discussion and reflection", (4) "students outcomes need greater consideration", (5) "students lack understanding on how to evaluate," (6) "personalized open-ended evaluations along with the BEST to get more in-depth information," (7) "more mentor/peer support is needed," and (8) "we cannot determine an effective system until we can articulate what is good teaching." One respondent indicated that student surveys were sufficient "once you have taught successfully for a period of time."

Literature regarding the use and effectiveness of student evaluations reveals that although many "faculty express cynicism about the accuracy of these instruments" (Beck, 2006 p. 18), a significant body of research suggests that course evaluation scores are the only practical indicator of teaching effectiveness with well-established validity of use (d'Apollonia & Abrami, 1997; Greenwald, 1997; Marsh & Roche, 1997; McKeachie, 1997). A few faculty members seemed to acknowledge an awareness of the correlation between BEST items and student learning. "Two of the global items on the BEST do show moderate relationship to student achievement," stated one faculty member at a departmental meeting. He referred to

a recent “study at IU” that found “if students report that their teachers are outstanding teachers they are 3x more likely to be engaged successfully and 5x more likely to be rated high by their professor that they had a high level of achievement in the class.” An awareness of the relationship between BEST item scores and student learning did not appear, however, to be common knowledge or an accepted belief of the majority of faculty present at the departmental meetings or responding to the electronic survey.

Promotion and Tenure/Annual Review

Themes appearing in the departmental meeting responses showed more dissatisfaction with the way teaching was evaluated for the purpose of annual merit review than the way it was evaluated for tenure and promotion. In addition, over 50% of the faculty and associate instructor respondents to the electronic survey were satisfied or mostly satisfied with the teaching evaluation instruments and documents used to assess the quality of teaching for tenure and promotion. Respondents described the promotion and tenure process as a “more well rounded” and “fair” process that utilized a variety of techniques and data for evaluation. This opinion was reiterated by the responses of departmental chairpersons. For example, one described the teaching evaluation process for promotion and tenure as “consistent and satisfactory [because] it genuinely requires multiple types of evidence.”

Dissatisfaction with departmental approaches to the evaluation of teaching seemed to focus on inconsistencies and vagaries implicit in merit review as a norm-based process. Faculty expressed a need for the articulation of differences between annual review data and tenure and promotion data. They communicated confusion regarding what to include as evidence of excellence in teaching for merit review. Faculty also wondered about the weight given to various data submitted for review, such as student evaluations, syllabi, unsolicited notes, etc.

Asked to comment on the value of the BEST instrument as a normative evaluation tool, one faculty member wrote:

There is a substantial literature on normative evaluation I have heard criticisms of the normative instrument for years usually by individuals who have low scores. While the validity research nationally would argue for continued use. The just criticism of the instrument is: there isn't very much variance at the top of the scale. So it is very difficult to differentiate between average and good performers. I believe the instrument is still quite useful in identifying extremely poor and extremely good performers (XXX, personal correspondence, January 29, 2009).

This faculty member's statement seems to agree with research that finds student ratings may identify those at extreme ends of a continuum of 'good teaching' (Berk, 2006), but they may not provide the kinds of nuanced information that would assist instructors in improving their pedagogic skills.

Retention of Associate Instructors

Neither faculty nor departmental chairpersons saw teaching evaluations as a dominant concern with respect to the retention of associate instructors, although a few faculty members communicated anecdotal concerns that associate instructors might "succumb to student pressure. Tougher graders are penalized if others are grading easier," and that low "BEST scores could haunt someone teaching a course for the first time." Associate instructors, however, did worry about their evaluations. Some believed these evaluations were of little value or were counter-productive to developing proficiency in teaching. Their responses to the electronic survey revealed a desire for more guidance and mentoring. For example, "I would like to see some kind of expectation for peer observations. As a classroom teacher this was extremely helpful and I believe it could be useful to guide discussion toward improving techniques and understanding our students' needs." One department was praised for offering support for AIs through program area meetings and monthly meetings during which ideas and resources for teaching were shared.

Concerns of Students

Participants of the Dean's Advisory Council focus group focused attention on the effectiveness of the BEST instrument as a tool for assessing the effectiveness of their instructors and instruction. All but

one of the participants (94%), believed that the BEST does not adequately evaluate the teaching effectiveness of faculty and associate instructors or accurately reflect learning. Reasons cited for this overwhelming agreement were the non-specificity of the items, and the recognition that BEST questions are not personalized to the instructor. “No, because I think the questions asked do not always apply to the specific teacher or course”, stated one respondent. “Assessment is very general”, said another. The lack of time permitted for writing a reflective analysis of course instructors and instruction was also commented upon in statements such as, “At the end of the period on the last day, no one wants to fill this “extra” stuff out”, and “I feel like I’m rushed to complete right after exams.”

What Changes If Any Would You Like to See?

Discussions at departmental meetings yielded a variety of suggestions for positive changes. These included “clearer guidelines for the consideration of teaching documents in terms of the annual review process”, “grouping courses into categories and evaluating them in comparison to others of that group”, and “summative observations by neutral expert colleagues trained to evaluate teaching.” A few (10%) of the respondents on the electronic survey sent to faculty and associate instructors suggested that the focus should be on learning outcomes, or that questions on the BEST instrument should inform what to improve through more open ended questions and question items. Departmental chairpersons agreed that guidelines in determining which items to include as evidence of good teaching would be helpful to both faculty and reviewers when making decisions about merit. Other changes, such as grouping courses into categories, seem reasonable in light of findings reported in the literature about the evaluation of teaching. Analyses of student ratings have shown “in general, students give higher ratings to elective courses, higher-level courses and courses in the humanities: the lowest ratings are reserved for required, lower-level survey courses and quantitative courses in math and statistics” (Berk, 2006, p. 200). These findings lend support to the wisdom of grouping courses into categories for the purpose of making comparisons. Also, tailoring BEST questions to the specifics of course content and instruction might be of benefit to instructors desiring

to improve teaching effectiveness. The request for observations by peer experts for the purpose of summative evaluation might be problematic, as one faculty member suggested, who wrote that in his experience (at least insofar as evaluating excellence of teaching for *summative* purposes is concerned), “Peer evaluation is notoriously biased” (XXX, personal correspondence, January 29, 2009). Several faculty and associate instructor respondents to the electronic survey (19%) indicated that they would like to see the teaching evaluation system introduce peer observation with informative feedback as a way of encouraging and supporting improvement in teaching. Associate instructors, particularly, wanted more time and opportunities to observe each other. “I am part of a very talented pool of people,” wrote one AI, “but unfortunately, I don't have the opportunity to observe and learn from them.” Overall, associate instructors seemed to believe (and a few experienced) observation and feedback as an aid in learning strategies of good teaching. This notion is supported by literature indicating that peer reviews, as alternative sources of evidence, are more appropriate as “formative rather than summative assessments” (Beck, 2006, p. 21)

Obstacles and Overcoming Obstacles

Determining effective instruments and approaches for determining excellence seemed particularly problematic with respect to the evaluation of teaching for merit review, as was evidenced in the statement, “Only a certain percentage can get an excellent rating – so we already start out with a limitation.” Regular faculty saw this as working against fostering an environment of learning within community.” Clinical faculty described an additional concern in that, “Clinical faculty practices are different from other professors, which impacts evaluations.” Overall, however, faculty voiced concern for achieving a balance between efficiency and quality evaluations. Several faculty worried that a new system based on more comprehensive data might not serve to change current problems or improve teaching abilities any better than the current system and might be more cumbersome and time consuming. “The obstacle is time,” wrote one faculty respondent, “We all have good ideas for evaluating teaching but we can't afford to spend more time monitoring teaching -- that time is better spent doing a good job of teaching.”

A solution to the obstacle of BEST as both the dominant instrument of evaluation and the least time-consuming form of evaluation might be that instructors “choose which BEST items to use” in order that these address specifics of course content and instruction. Faculty also suggested the need for “a rubric for merit review” or detailed guidelines about what to include in merit reviews with explanations about how documents and evidence are to be weighted. One faculty member added, “First year evaluations should not be weighed so heavily.” The need for guidelines was reiterated by associate instructors. One wrote, “Clearly explain to AIs how they will be evaluated and why.” Two respondents to the electronic survey sent to faculty and associate instructors provided caveats regarding recommendations for change, when asked if there was “anything else you would like to add or comment upon.”

Some types of evidence probably do deserve greater weight than others; for example, student eval[uations] do seem more important, when taken as a whole . . . compared with the good intentions of teaching, such as elaborate syllabi. . . . [E]vidence of student achievement or student products tell us little or nothing about student GROWTH/LEARNING as a result of a particular class, and I would place little weight on "presentation or articles about one's teaching". How I describe or interpret my own teaching can be quite far afield from what matters most, which is how students experience our teaching.

In all levels of education, teachers are inundated with excess work in the name of accountability that limits time spent on actual teaching. If you are looking to change the evaluation procedures of teachers in this department, I hope there is a valid reason and that the procedures are reasonable requests rather than arbitrarily determined by a board that will be dissolved before

actually having to see the implementation of these ideas through to their conclusion.

Summary

Findings show that regular faculty and departmental chairpersons generally saw some consistency with the statement from the *Promotion and Tenure Criteria for the School of Education* (02.30R, p. 4) (Appendix 1) and the current system of evaluation teaching insofar as promotion and tenure are concerned, and were generally satisfied in this regard. However, they saw inconsistencies and were dissatisfied with the application of this system insofar as annual merit reviews were concerned. In fact, a majority of faculty and associate instructors responding to the electronic survey expressed dissatisfaction or mixed feelings with their department's evaluation of teaching. There was a pervasive sense of confusion about what to provide for the annual merit review. Faculty communicated that annual merit review guidelines should provide clarity regarding what evidence of quality teaching to include and how various kinds of data will be weighted. They offered a number of specific suggestions to accomplish this objective. Associate instructors conveyed a need for articulation of what quality teaching is, and a desire for more guidance and mentoring to assist them in becoming good teachers. Students, and to a lesser degree instructors, suggested specific ways to better utilize BEST to elicit evaluation data better aligned with course content and instruction. Finally, all the queried stakeholders in the teaching evaluation process seemed to agree that evaluations, in and of themselves, do not improve teaching. Approaches are needed that foster a culture of continual improvement.

Recommendations

Based on our findings, we propose the following recommendations for consideration by the Committee on Teaching. We request that the Committee on Teaching forward our report to Policy Council for possible action.

- Articulate the distinction between evaluating teaching for promotion and tenure and for

annual merit review, with respect to the purpose of evaluation as well as the criteria by which an evaluative judgment will be made. The two processes should be made more transparent. Our data also raises the question whether multiple forms of evidence (a more time-intensive evaluative process) should be required for annual merit review as they are now for promotion and tenure.

- Establish common criteria across departments within the School of Education regarding materials required for submission as evidence of quality teaching *and* their weighting.
- Group courses in categories when using student ratings as evaluative data: for example, compare introductory courses with other introductory courses, graduate seminars with other graduate seminars, online courses with online courses, and so forth.
- Delineate the weight given to various forms of documentation.
- Assign less weight to 1st year evaluations.
- Design and implement a coherent and consistent program of guidance and mentoring for associate instructors that is ongoing throughout the academic year. Given the amount of concern about this issue voiced in our data, it seems that the current beginning- of- the- year orientation provided by the Office of Teacher Education is not sufficient. Clearly, associate instructors are requesting support and opportunities to improve their teaching skills.
- Inform associate instructors as to the purpose of evaluating their teaching, the method of doing so, and how these data will be utilized for the purpose of an evaluative decision.
- Clarify to all course instructors the role they can play in constructing the course evaluation instrument (i.e., that they have a choice in selecting which items appear on the

BEST evaluation form).

- Advise new instructors to more carefully match BEST items with specific course content and instruction.
- If course evaluations are administered in class, allow adequate time for students to provide thoughtful responses to BEST questions, pending school-wide adoption of online administration of BEST.

Policy Statements Based on Recommendations of the Report

Recommendation 1: Articulate the distinction between evaluating teaching for promotion

and tenure and for annual merit review, with respect to the purpose of evaluation as well as the criteria by which an evaluative judgment will be made. The two processes should be made more transparent. Our data also raises the question whether multiple forms of evidence (a more time-intensive evaluative process) should be required for annual merit review as they are now for promotion and tenure.

The evaluation of teaching for the purpose of promotion and tenure and for annual merit review differs. Both forms of evaluation are summative in nature. Evaluation for the purpose of promotion and tenure, however, is criterion-based and for annual merit review norm-based. In the first case, the candidate is compared with an acknowledged standard of excellence and in the second case, faculty are compared with one another.

The *Promotion and Tenure Criteria for the School of Education* appears to acknowledge that teaching is a complicated and context-bound process. Consequently, the evaluation of teaching also might be complex and multiple sources of data might be utilized in determining effective teaching. This recommendation currently guides preparation of the teaching component of the dossier. In annual merit review, a tradition of heavy reliance on student ratings of instruction has gained acceptance. There appears

to be a clear consensus among various stakeholders in the School of Education that the evaluation of teaching for annual merit review not be a more time-intensive process than currently exists.

For the purpose of annual merit review, student ratings will continue to be required. Researchers in the field (e.g., d'Apollonia & Abrami, 1997; Greenwald, 1997; Marsh & Roche, 1997; McKeachie, 1997) agree that course ratings are an indicator of teaching effectiveness with well-established *validity* of use (Kulik, 2001; Scriven, 1981). Nonetheless, skepticism remains about the value of student ratings of teaching in spite of their documented validity of use (Berk, 2006). According to Arreola (2007, p. 98), “faculty and administrators continue to ask questions concerning the factors that may, or may not, influence student ratings of instructors and instruction.” Given this lack of confidence, evaluative evidence also should include additional data. If additional data are submitted, this evidence *must* be taken into consideration. The option of doing so acknowledges the complexity involved in evaluating teaching and makes it better aligned with the mode of evaluation in place for promotion and tenure. (See *Promotion and Tenure Criteria for the School of Education*, 02.30R, p. 4. Retrieved May 29, 2009, from <http://site.educ.indiana.edu/Portals/28/Policy%20Council/2001-2002/02.30R.pdf>)

Recommendation 2: Establish common criteria across departments within the School of Education regarding materials required for submission as evidence of quality teaching *and* their weighting.

The criteria employed in the evaluation of teaching should be the same across all five departments in the School of Education. Weighting of these criteria should also be similar among departments. The rubric shown below specifies criteria and weighing of these criteria that should be used by those responsible for evaluating faculty performance for the purpose of annual merit review. (See rubric).

Additional guidelines include:

- Every department in the School of Education should formulate a statement informing their associate instructors as to the purpose of evaluating their teaching, the method of doing

so, and how these data will be utilized. The statement should be submitted for approval by the Committee on Teaching by January 1, 2010.

- At the beginning of each academic year, every course coordinator should make transparent the role an associate instructor can play in constructing the course evaluation instrument (i.e., whether or not he or she has a choice in selecting which items appear on the BEST evaluation form). New instructors should be advised about how to carefully match BEST items with specific course content and instruction. In addition, pending universal adoption of on-line student ratings of teaching, instructors should be instructed to allow adequate time for students to provide thoughtful responses to BEST questions. This information should be disseminated verbally and in writing by the Associate Dean of Teacher Education and the Associate Dean for Graduate Studies (e.g., via the *Teacher Education Instructor Resources Handbook* disseminated by the Office of Teacher Education).

Recommendation 3: Design and implement a coherent and consistent program of guidance and

mentoring for associate instructors that is ongoing throughout the academic year. Given the amount of concern about this issue voiced in our data, it seems that the current beginning-of-the-year orientation provided by the Office of Teacher Education is not sufficient. Clearly, associate instructors are requesting support and opportunities to improve their teaching skills.

Each department in the School of Education should take steps to set up a “Teaching Development Community” for faculty and assistant instructors. Based on the premise that the improvement of teaching is an ongoing, lifelong process, its aim would be to provide a non-threatening and supportive environment to facilitate professional growth. The provision of a group setting where faculty and associate instructors interact collegially on an equal footing would be a distinguishing characteristic of the TDC. For departments currently without an established mechanism for the development of teaching, the TDC would address what appears to be an issue of concern voiced by some respondents. For other departments that

have an ongoing program of guidance and mentoring for associate instructors, the TDC could be incorporated into what is already in place.

Suggested activities might include: (1) sharing of “best practices,” syllabi, instructional materials and strategies, knowledge of effective instructor and student dispositions, and so forth; (2) linking of participants with available campus-wide resources on teaching (e.g., Disability Services, the Student Academic Center); (3) taking advantage of Campus Instructional Consulting services; (4) networking with faculty and associate instructors recognized for excellence in teaching; and (5) attending events sponsored by the Scholarship of Teaching and Learning.

Given the amount of concern voiced in our data about the need for a coherent program of support in spite of the support that does currently exist in the School of Education, one cannot help but ponder why respondents appeared to be asking for something more. The support that currently exists may be underutilized because of a lack of awareness of the value of the expertise offered by Instructional Consulting Services (ICS) or insufficient time to take advantage of the available resources. Additionally, those needing ICS may be embarrassed, fearful of recrimination (i.e., loss of work), and mindful of a “stigma” (real or perceived) associated with seeking out assistance. Designed as a collegial forum attended by faculty and associate instructors alike with varying levels of expertise in teaching, the TDC is intended to overcome these barriers.

Each course coordinator should submit each semester a plan of guidance and mentoring to the department chairperson. The Associate Dean of Teacher Education should send out a reminder of this policy requirement to department heads each fall and spring. Department chairpersons should collect and forward these plans to the Office of Teacher Education, where they should be kept on file.

School-wide Rubric for Rating the Evaluation of Teaching

| Evidence | Exemplary | Meritorious | Satisfactory | Unsatisfactory |
|---|-----------|-------------|--------------|----------------|
| <p>1. Student Ratings</p> <p>Required Questions:</p> <ul style="list-style-type: none"> ○ Overall, I would rate the quality of this course as <ul style="list-style-type: none"> ○ outstanding ○ good ○ satisfactory ○ unsatisfactory ○ Overall, I would rate this instructor as <ul style="list-style-type: none"> ○ outstanding ○ good ○ satisfactory ○ unsatisfactory ○ In general it seemed that the instructor was well-prepared for class sessions. ○ From my perspective the instructor respected and showed interest in his/her students and their ideas. ○ In general the atmosphere of this class supported students learning. <p>Recommended Questions:</p> <ul style="list-style-type: none"> ○ My instructor uses teaching strategies that encourage me to engage in higher-order thinking (e.g., apply, problem solve, question, evaluate). ○ The writing assignments in this course encourage me to engage in higher-order thinking (e.g., apply, problem solve, question, evaluate). <p>Customized Questions:</p> <ul style="list-style-type: none"> ○ Students' responses to customized questions of the instructors' choice | | | | |
| <p>2. Syllabi</p> <ul style="list-style-type: none"> ○ Contents should include course description, schedule of topics, objectives, required readings, assignments and their weighting (grading criteria). ○ Are the contents of the syllabi detailed, thoughtful, and well articulated? | | | | |
| <p>3. Scholarship of Teaching</p> <ul style="list-style-type: none"> ○ Publications and teaching-related projects. Teaching-related projects should be interpreted broadly as funded teaching projects, piloting in anticipation of funding for teaching projects, course development (regardless of whether courses make), systematic participation in teaching development activities (e.g., SOTL, TDC events, TLTC Workshops, etc.), willingness to direct independent studies, and so forth. ○ Service-related teaching outside the confines of a traditional classroom (e.g., podcasting, blogging, video conferences, etc.). ○ Open teaching activities (e.g., instructional delivery to audiences around the world, and other teaching activities for which students are not formally registered for grades). <p>It is incumbent upon those submitting materials to provide documentation of the quality of these activities.</p> | | | | |
| <p>4. Other Forms of Evidence</p> <ul style="list-style-type: none"> ○ Philosophy statement ○ Reflection on efforts to evaluate and improve teaching ○ Observations by colleagues ○ Teaching awards and nominations for awards | | | | |

| | |
|--|--|
| <p>Whenever additional data are submitted beyond student ratings, they should be weighted according to the extent to which they provide evidence of competent teaching</p> | |
| <p>Certain factors ought to be taken into consideration when evaluating student ratings.</p> <ul style="list-style-type: none"> ○ Student comments may sometimes indicate that certain factors call into question the validity of students' judgments about the course and/or teaching: e.g., course rigor (theory vs. practice, grades), technological difficulties with Oncourse, group dynamics, and so forth. <i>Rate more generously.</i> ○ Introductory level courses or newly developed courses typically receive lower ratings than upper level and graduate courses. <i>Rate more leniently.</i> ○ Atypical field experiences, and online or distance learning courses require more instructional time than other courses. <i>Rate more generously.</i> | <p><i>(write in special considerations & indicate how these were considered)</i></p> |

References

- Abrami, P., d'Apollonia, S., Cohen, P. (1990). Validity of student ratings of instruction: What we know and what we do not. *Journal of Educational Psychology, 82*(2), 219-231.
- Beck, R. A. (2006). *Thirteen strategies to measure college teaching*. Sterling, VA: Stylus.
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research, 51* (3), 281-309.
- d'Apollonia, S., & Abrami, P.C. (1997). Navigating student ratings of instruction. *American Psychologist, 52*, 1198-1208.
- Doyle, K. O., Jr. (1983). *Evaluating teaching*. Lexington, MA: Lexington Books.
- Feldman, K. A. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education, 30*, 583–645.
- Greenwald, A.G. (1997). Validity concerns and the usefulness of student ratings of instruction. *American Psychologist, 52*, 1182-1186.
- Kulik, J. A. (2001). Student ratings: Validity, utility and controversy. *New Directions for Institutional Research, 109*, 9-25.
- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76*(5), 707–754.
- Marsh, H. W. (1987). Students' evaluation of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11*, 253-387.
- Marsh, H. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity,

- potential biases and usefulness. In R. Perry & J. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319-383). Dordrecht, The Netherlands: Springer.
- Marsh, H.W., & Roche, L. (1994). *The use of students' evaluations of university teaching to improve teaching effectiveness*. Canberra, Australia: Australian Government Publishing Service.
- McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist*, 52, 1118 -1225.
- Promotion and Tenure Criteria for the School of Education*, 02.30R. Retrieved February 23, 2009, from <http://site.educ.indiana.edu/Portals/28/Policy%20Council/2001-2002/02.30R.pdf>
- Scriven, M. (1981). Summative teacher evaluation. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp. 244-271). Beverly Hills, CA: Sage.
- Scriven, M. (1989). The design and use of forms for the student evaluation of teaching. *Instructional Evaluation*, 10, 1-13.
- Spencer, P. A., & Flyr, M. L. (1992). *The formal evaluation as an impetus to classroom change: Myth or reality?* (ERIC Document Reproduction Service No. ED349053).

Appendix

Table of Contents

1. Electronic Survey for Faculty and Associate Instructors
2. Departmental Focus Group Survey
3. Interview Questions for Departmental Chairpersons
4. Student Focus Group (Dean's Advisory Council) Survey
5. Data from Electronic Survey for Faculty and Associate Instructors
6. Data from Departmental Focus Group Survey
7. Data from Departmental Chairperson Interviews
8. Data from Student Focus Group (Dean's Advisory Council) Survey

**School of Education Committee on Teaching Subcommittee
to Review the Evaluation of Teaching**

Electronic Survey for Faculty and Associate Instructors

In the School of Education, evidence of excellence in teaching is collected for a number of purposes. For example, the stated policy regarding the evaluation of teaching in regards to tenure and promotion is stated as:

Teaching is a complex process that encompasses multiple components, *multiple forms of evidence are needed to assess teaching effectiveness comprehensively*. No one form of evidence (e.g., student evaluations of teaching) should carry disproportionately more weight than any other. *Appropriate* teaching materials may include evidence from the *instructor* (e.g., philosophy of teaching, teaching goals, syllabi, instructional materials, reflections on efforts to evaluate and improve teaching, presentations and articles on one's teaching), evidence from *others* (e.g., colleague evaluations of student outcomes, observations by colleagues trained to evaluate teaching, invitations to share one's teaching expertise), and evidence from *students* (e.g., solicited and unsolicited feedback from students, course-related student products, evidence of student achievement, student-selected teaching awards). These categories of evidence may be interrelated. For example, a colleague may write an evaluation of the links among an instructor's philosophy, goals, course design, instructional strategies, and outcomes based on direct observation, instructor-provided documents, and student products and evaluations.

(taken verbatim from *Promotion and Tenure Criteria for the School of Education*, 02.30R, p. 3. Available online at:

<http://site.educ.indiana.edu/Portals/28/Policy%20Council/2001-2002/02.30R.pdf>)

The purpose of this survey is to determine how faculty and associate instructors perceive current practices related to this policy and the extent to which they are satisfied or dissatisfied with these practices. Please respond to these questions:

In your estimation is the way teaching is evaluated in your department consistent with this statement?

If not, how does it differ?

How is your teaching evaluated in your department? Are there multiple kinds of assessment data (e.g., the BEST survey, peer observations)? Please describe.

To what extent are you satisfied with the approach to evaluating teaching in your department? Please check the extent to which you are satisfied (dissatisfied) with respect to each of these areas. Then, please comment.

Merit review for faculty

| completely satisfied | mostly satisfied | mixed (both satisfied/dissatisfied) | mostly dissatisfied | completely dissatisfied | not applicable |
|----------------------|------------------|-------------------------------------|---------------------|-------------------------|----------------|
| | | | | | |

Comments

Promotion and tenure for faculty

| completely satisfied | mostly satisfied | mixed (both satisfied/dissatisfied) | mostly dissatisfied | completely dissatisfied | not applicable |
|----------------------|------------------|-------------------------------------|---------------------|-------------------------|----------------|
| | | | | | |

Comments

Hiring and retention of associate instructors

| completely satisfied | mostly satisfied | mixed (both satisfied/dissatisfied) | mostly dissatisfied | completely dissatisfied | not applicable |
|----------------------|------------------|-------------------------------------|---------------------|-------------------------|----------------|
| | | | | | |

Comments

What changes, if any, would you like to see in the way teaching is currently being evaluated in your department?

What obstacles, if any, do you perceive in revising the way teaching is currently evaluated in your department?

How might these obstacles be overcome or resolved?

Is there anything else you would like to add or comment upon regarding the evaluations or the evaluation process?

Please check your position title as shown below.

Departmental Affiliation ___CEP ___C&I___ELPS ___IST___LCLE

_____Years teaching in the IUB School of Education (total)

_____ Regular/Clinical Faculty Member

_____Assistant Professor

_____Associate Professor

_____Full Professor

_____Adjunct

_____Associate Instructor

**School of Education Committee on Teaching Subcommittee
to Review the Evaluation of Teaching**

Departmental Focus Group Survey

In the School of Education, evidence of excellence in teaching is collected for a number of purposes. For example, the stated policy regarding the evaluation of teaching in regards to tenure and promotion is stated as:

Teaching is a complex process that encompasses multiple components, *multiple forms of evidence are needed to assess teaching effectiveness comprehensively*. No one form of evidence (e.g., student evaluations of teaching) should carry disproportionately more weight than any other. *Appropriate* teaching materials may include evidence from the *instructor* (e.g., philosophy of teaching, teaching goals, syllabi, instructional materials, reflections on efforts to evaluate and improve teaching, presentations and articles on one's teaching), evidence from *others* (e.g., colleague evaluations of student outcomes, observations by colleagues trained to evaluate teaching, invitations to share one's teaching expertise), and evidence from *students* (e.g., solicited and unsolicited feedback from students, course-related student products, evidence of student achievement, student-selected teaching awards). These categories of evidence may be interrelated. For example, a colleague may write an evaluation of the links among an instructor's philosophy, goals, course design, instructional strategies, and outcomes based on direct observation, instructor-provided documents, and student products and evaluations.

(taken verbatim from *Promotion and Tenure Criteria for the School of Education*, 02.30R, p. 3. Available online at:

<http://site.educ.indiana.edu/Portals/28/Policy%20Council/2001-2002/02.30R.pdf>)

The purpose of this interview is to determine how faculty and associate instructors perceive current practices related to this policy and the extent to which they are satisfied or dissatisfied with these practices. Please respond to these questions:

1. In your estimation is the way teaching is evaluated in your department consistent with this statement?

2. If not, how does it differ?

3. To what extent are you satisfied with the approach to evaluating teaching in your department? Please comment about the extent of your satisfaction with respect to each of these areas: (1)

merit review for faculty; (2) tenure and promotion of faculty; and, (3) hiring and retention of associate instructors.

4. What changes, if any, would you like to see in the way teaching is currently being evaluated in your department?
5. What obstacles, if any, do you perceive in revising the way teaching is currently evaluated in your department?
6. How might these obstacles be overcome or resolved?
7. Is there anything else you would like to add or comment upon regarding the evaluations or the evaluation process?

**School of Education Committee on Teaching Subcommittee
to Review the Evaluation of Teaching**

Interview Questions for Departmental Chairpersons

In the School of Education, evidence of excellence in teaching is collected for a number of purposes. For example, the stated policy regarding the evaluation of teaching in regards to tenure and promotion is stated as:

Teaching is a complex process that encompasses multiple components, *multiple forms of evidence are needed to assess teaching effectiveness comprehensively*. No one form of evidence (e.g., student evaluations of teaching) should carry disproportionately more weight than any other. *Appropriate* teaching materials may include evidence from the *instructor* (e.g., philosophy of teaching, teaching goals, syllabi, instructional materials, reflections on efforts to evaluate and improve teaching, presentations and articles on one's teaching), evidence from *others* (e.g., colleague evaluations of student outcomes, observations by colleagues trained to evaluate teaching, invitations to share one's teaching expertise), and evidence from *students* (e.g., solicited and unsolicited feedback from students, course-related student products, evidence of student achievement, student-selected teaching awards). These categories of evidence may be interrelated. For example, a colleague may write an evaluation of the links among an instructor's philosophy, goals, course design, instructional strategies, and outcomes based on direct observation, instructor-provided documents, and student products and evaluations.

(taken verbatim from *Promotion and Tenure Criteria for the School of Education*, 02.30R, p. 3. Available online at:
<http://site.educ.indiana.edu/Portals/28/Policy%20Council/2001-2002/02.30R.pdf>)

The purpose of this interview is to determine how departmental chairpersons perceive current practices related to this policy and the extent to which they are satisfied or dissatisfied with these practices. Please respond to these questions:

In your estimation is the way teaching is evaluated in your department consistent with this statement?

If not, how does it differ?

How is teaching evaluated in your department? Are multiple kinds of assessment data (e.g., the BEST survey, peer observations) collected? Please describe.

To what extent are you satisfied with the approach to evaluating teaching that you described? Please comment about the extent of your satisfaction with respect to each of these areas: (1) merit review for faculty, (2) promotions and tenure for faculty, (3) hiring and retention of associate instructors?

In your estimation, how satisfied are the faculty in your department with respect to each of these areas: (1) merit review for faculty, (2) promotions and tenure for faculty, (3) hiring and retention of associate instructors?

What obstacles, if any, do you perceive in revising the way teaching is currently being evaluated in your department?

How might these obstacles be overcome or resolved?

Is there anything else you would like to add or comment upon regarding the evaluations or the evaluation

process?

**School of Education Committee on Teaching Subcommittee
to Review the Evaluation of Teaching**

Student Focus Group (The Dean's Advisory Council) Survey

1. Does the BEST course evaluation instrument (the standard computer-scanned survey), as it exists, adequately evaluate the teaching effectiveness of faculty and associate instructors?

2. Does the BEST instrument give you an adequate opportunity to express your learning? Please explain.

3. Does the BEST instrument give you an adequate opportunity to express your estimation of an instructor's teaching? Please explain.

4. If not, what alternate means of evaluating would you suggest?

5. Would your peers agree that the BEST course evaluation instrument, as it exists, provides them an opportunity to express their estimation of their learning? Please explain.

6. Would your peers agree that the BEST course evaluation instrument, as it exists, provides them an opportunity to express their estimation of their instructor's teaching? Please explain.

Data from Electronic Survey for Faculty and Associate Instructors

1/30/09

Wendy Wall-Marencik and Rajat Chadha

Demographic Information

Total Respondents (N=51)

Faculty Status Information

Associate Instructors AI (N=26)

Regular Faculty Members RFM (N=20)

Clinical Faculty Members CFM (N=2)

No Response (N=3)

Department Representation

Curriculum and Instruction (N=20)

Educational Leadership and Policy Studies (N=10)

Counseling and Educational Psychology (N=7)

Literacy, Culture and Language Education (N=7)

Instructional Systems Technology (N=6)

Missing: (N=1)

AI teaching yrs

<6mths (N=3)

< 1yr (N=0)

1 - 2 yrs (N=8)

2 - 3 yrs (N=6)

3 - 4 yrs (N=7)

4 - 5 yrs (N=1)

>5 yrs. (N=2)

Total (N=27)

Survey Questions 1-8

1. In your estimation is the way teaching is evaluated in your department consistent with this statement?

Responses:

| Faculty Status | “No” Response | “Yes” Response | Other Response* |
|-----------------------|----------------------|-----------------------|------------------------|
| Total (N=50) | 26 | 12 | 12 |
| RFM (N=21) | 10 | 7 | 4 |
| CFM (N=2) | 1 | 1 | 0 |
| AI (N=25) | 14 | 3 | 8 |
| Unknown (N=2) | 1 | 1 | |

*Response too vague for either a yes or no (e.g., “somewhat”, “not sure”)
(missing data 1)

Summary and Themes

Out of the 50 respondents on this item, 21 (42%) are regular faculty members (RFM), 2 are clinical faculty members (CFM), 25 (50%) are associate instructors (AI) and 2 are of unknown status. Of the 21 RFM, approximately 50% responded “no” they did not estimate the way teaching is evaluated in their department was consistent with the introductory statement. Of these RFMs, approximately 4% responded “yes” they did believe teaching evaluation was consistent with the statement. Approximately 19% of the RFM were vague or mixed responses. Of the only 2 clinical faculty members (CFM) who responded, one replied “yes” and the other “no”. Associate instructors were the highest number of respondents with approximately 50% indicating a “yes”, 12% “no” and 32% responded with vague or nondefinitive responses. Two respondents selected “yes”, but did not complete the status information.

Many of the individuals, that responded “yes” also listed some stipulations with their answer, for example, many qualified “yes, for the most part” or “by in large” or “my experience is limited”. The “nos” tended to be more definitive in their responses. The majority of “no” responses focused on the use of “only the student evaluation” as a measure. Also among the “nos” were strong statements such as the following from as CFM and AI, “No, it is not, I have never been observed, never been given feedback, and never had someone explain to me the efficacy of the different percentages on the bubble survey (BEST)” “Absolutely not. The student evaluations are used excessively, despite being misleading in and of themselves”. The main themes identified within the “other-too vague” respondents focused on comments of “best intentions and attempts” to meet these standards, or other methods are in place but are discounted

or valued less. Question 2 gets more to the heart of the issues.

2. In not, how does it differ?

Summary and Themes

Out of the 38 respondents to this question, 35 report that student evaluation is the primary source of data. Some comments followed with specific criticisms, “Our dept does a poor job of communication with its AIs about their performance in the classroom”, “Focus is placed on evaluations with very little attention paid to actual class and the events within that space.” “Student evaluations carry disproportionately more weight”. Three respondents identified the use of syllabi and one identified the use of essays in the application process as part of their teaching evaluation. One respondent with CEP shared that student evaluation, philosophy statements, self informal evaluation and syllabi are all used for teacher evaluation.

3. How is your teaching evaluated in your department? Are there multiple kinds of assessment data (e.g., the BEST survey, peer observations)? Please describe.

Responses:

| Types of Assessments Used | Number out of N=47 Respondents |
|-------------------------------|--------------------------------|
| Student survey(BEST) | 47 |
| Peer observation | 13 |
| Mentor Observation | 5 |
| Syllabi | 3 |
| Personalevaluation/statements | 3 |
| Student letters | 1 |
| Course meetings | 1 |
| Survey monkey | 1 |
| Widerange (non specific) | 1 |
| Publication | 1 |

(missing data 5)

Summary and Themes:

All 47 respondents identified student survey as required and the primary method for evaluation. For promotion and tenure, other methods such as peer reviews, syllabi, student letters, and personal statements are required by some departments. For annual reviews for faculty, some departments require other documentation to be submitted, such as syllabi. AIs required to complete portfolios, share that personal statements, syllabi, mentor observation are also required for teacher evaluation. According to the AI responses, most mentor observations are limited to one. In several departments, peer observations are encouraged but not required for faculty.

4. To what extent are you satisfied with the approach to evaluating teaching in your department? Please check the extent to which you are satisfied (dissatisfied) with respect to each of these areas. Then please comment.

A. For merit review for faculty

Responses:

| Completely Dissatisfied | Mostly Dissatisfied | Mixed | Mostly Satisfied | Completely Satisfied | Not applicable |
|-------------------------|---------------------|-------|------------------|----------------------|----------------|
| 4 | 7 | 16 | 9 | 8 | 6 |

(missing data 1)

Summary and Themes:

Out of N=44 respondents, excluding 6 not applicable, 4 (9%) are completely dissatisfied with the approach their department employs for evaluating teaching for merit review for faculty, 7 respondents; approximately 18% are mostly dissatisfied. The highest response was 16 representing mixed feelings about their department's evaluation process at approximately 36%. Of the respondents 9 (nearly 21%) are mostly satisfied and 8 respondents (approximately 18%) are completely satisfied.

Many of the comments in support of the student survey identified issues with "efficiency" and believe student surveys are sufficient once you have taught successfully for a period of time. For those with mixed feelings and dissatisfaction identified the following issues as weaknesses for their current systems; 1) more flexibility is needed, 2) evaluations are ignored, useless, and do not improve teaching skills, 3) more formal time needs to be structured for discussion and reflection, 4) students outcomes need greater consideration (i.e., achievement, work), 5) we cannot determine an effective system until we can articulate what is good teaching, 6) personalized open-ended evaluations along with the BEST to get more in-depth information, 7) too much emphasis on this one assessment, 8) more mentor/peer support is needed, and 9) students lack understanding on how to evaluate. Concerns were also expressed by several individuals with mixed responses that a new system may not serve to change the current problems with ignoring and improving teaching abilities based on more comprehensive data any better than the current system and may be more cumbersome and time consuming. Some voice concern for a balance between efficiency and quality evaluations.

B. For promotion and tenure

Responses:

| Completely Dissatisfied | Mostly Dissatisfied | Mixed | Mostly Satisfied | Completely Satisfied | Not applicable |
|-------------------------|---------------------|-------|------------------|----------------------|----------------|
| 3 | 4 | 6 | 5 | 9 | 16 |

(missing data 8)

Summary and Themes:

Out of N=27 respondents, excluding 16 not applicable, 3 (11%) are completely dissatisfied with the approach their department employs for promotion and tenure, 4 respondents; approximately 15% are mostly dissatisfied. Those with mixed feelings about their department's evaluation process at were at approximately 22%. Of the respondents 5 (nearly 19%) are mostly satisfied and 9 respondents (approximately 33%) reflected highest number at completely satisfied with the process for promotion and tenure.

Based on the feedback, many respondents believed that the promotion and tenure process was a "more well rounded" and "fair" process considering a variety of techniques and data for evaluation. Some dissatisfied with the process felt their dossier was scarcely reviewed and the department relied more on comments not grounded in evidence.

C. For hiring and retention of associate instructors

Responses:

| | | | | | |
|-------------------------|---------------------|-------|------------------|----------------------|----------------|
| Completely Dissatisfied | Mostly Dissatisfied | Mixed | Mostly Satisfied | Completely Satisfied | Not applicable |
| 3 | 9 | 10 | 8 | 5 | 11 |

(missing data 4)

Out of N=35 respondents, excluding 11 not applicable, 3 (approximately 9%) are completely dissatisfied with the approach their department for hiring and retention of AIs, 9 respondents; approximately 26% are mostly dissatisfied. Those with mixed feelings about their department's evaluation process, 10 were at approximately 29%. Of the respondents 8 (nearly 23%) are mostly satisfied and 5 respondents (approximately 14%) reported being completely satisfied with the process for hiring and retention. The majority of respondents fell within categories indicating less assurance with their feelings, however this was a fairly even mix between satisfaction and dissatisfaction.

Perhaps the reason this group responded in more vague categories has to do the many narrative responses indicating the lack of knowledge on how hiring and retention of AIs is conducted or how the student evaluations are used for this purpose. Some respondents indicated that multiple measures are necessary in this process, such as formal observations and measurement of student outcomes. What is most evident based on many comments in this section is a clear concern for AIs who are not prepared to be teaching. Respondents believed AIs need more guidance and mentoring. Concern was indicated for undergraduate students who are taught by AI who are not prepared to teach. AIs worry about their evaluations and some believe these evaluations are of little value or become counter-productive to developing proficiency in teaching. Some departments (LCLE) offer support for AIs through program area meetings and monthly E & X meetings for sharing ideas and resources. Finally, it was noted that not all AIs get classroom experience.

5. What changes, if any, would you like to see in the way teaching is currently being evaluated in your department?

The change that most of the respondents (9) would like to see in the teaching evaluation system is to introduce peer observation with informative feedback to improve teaching. Few (4) respondents

| Change | No of respondents |
|------------------------------|--------------------------|
| Use peer observations | 9 |
| Focus on learning outcomes | 4 |
| What is good teaching? | 3 |
| Faster feedback from BEST | 2 |
| Already use multiple sources | 3 |

Some other suggestions that were mentioned rarely were: AIs should not be evaluated for course design if they are not responsible for it, different questions on BEST evaluation and some questions do not inform what to improve and more open ended questions, focus on improvement of items scores.

“I would like to see some kind of expectation for peer observations. As a classroom teacher this was extremely helpful and I believe it could be useful to guide discussion toward improving techniques and understanding our students' needs.”

“Two, allow time and resources for AIs to observe each other. I am part fo a very talented pool of people, but unfortunately, I don't have the opprotunity to observe and learn from them.”

“Peer observation & evaluations with questions more closely related to the content and structure of courses taught. The survey that is used now is too general and doesn't give specific feedback to help me as an

instructor make changes in future classes. I have my students fill out a more helpful evaluation that I've created that gives me the best feedback in planning for the future semester.”

“Student achievement and student work should be considered primarily, since this shows that students learned what they were supposed to. Other factors like students' perceptions of professionalism and skill should also be considered, but those should be secondary and qualitative in nature, not a rating scale”

6. What obstacles, if any, do you perceive in revising the way teaching is currently evaluated in your department?

15 respondents reported that the need for extra time is the biggest obstacle in changing the current evaluation system. Other obstacles reported rarely are importance of research in a faculty member's career at IU, resistance to change, criteria for good teaching, lack of resources.

“The obstacle is time. We all have good ideas for evaluating teaching but we can't afford to spend more time monitoring teaching -- that time is better spent doing a good job of teaching.”

Addendum:

This additional data was sent forward by several participants at a later time and provides this feedback

- 1) Generating more or alternate forms of 'proof' around our teaching, we will end up requiring people to do more work;
- 2) Extra evidence beyond the BEST scores will not be likely to affect the ranking in any case.
- 3) Such a requirement would result in more work for people on the merit review committee, making an onerous job worse.

“People are just too busy to have a thorough discussion of teaching.”

7. How might these obstacles be overcome or resolved?

There is no common theme that emerges from the responses. There are diverse responses to this question and the number of responses is also low (27 out of 50).

“We will always be pressed for time and it will take quite a bit of time for the culture of our department to shift, but we are taking steps forward, which is what we need to do.”

“Perhaps added support personnel to oversee and management teacher evaluation and collect outcome data which could serve two purposes 1)improved evaluation process 2)formative feedback.”

“Priorities could be changed, but I don't see it because research brings in grant money and this is the surest way for a professor to ensure tenure and longevity.”

“Provide departmental teaching workshops to facilitate the discussion of teaching among colleagues.”

“We should document the efforts we put into revising and updating our courses.”

Addendum:

Leave it up to the individual to add additional evidence if he/she thinks that the BEST scores do not reflect the quality of his/her teaching.

People always have the right to appeal if they believe that they have been unfairly ranked.

It might be fair to take into account that some courses are more difficult to teach than others and that some courses seem historically to get rather low BEST scores across most instructors.

8. Is there anything else you would like to add or comment upon regarding the evaluations or the evaluation process?

Some of the comments were:

“If a class is co-taught - there should be an evaluation form to fill out on both instructors - not just one form. Two people can (and usually do) have very different methods of instruction and beliefs on how they "conduct" class.”

“some types of evidence probably do deserve greater weight than others; for example, student evals do seem more important, when taken as a whole (not just one or two disenchanted students) compared with the good intentions of teaching, such as elaborate syllabi also, evidence of student achievement or student products tell us little or nothing about student GROWTH/LEARNING as a result of a particular class and I would place little weight on "presentation or articles about one's teaching"--how I describe or interpret my own teaching can be quite far afield from what matters most, which is how students experience our teaching”

“Clearly explain to AIs how they will be evaluated and why.”

“Make the formal student evaluation available as soon as possible.”

“In all levels of education, teachers are inundated with excess work in the name of accountability that limits time spent on actual teaching. If you are looking to change the evaluation procedures of teachers in this department, I hope there is a valid reason and that the procedures are reasonable requests rather than arbitrarily determined by a board that will be dissolved before actually having to see the implementation of these ideas through to their conclusion.”

Addendum Data:

1. The old distinction between normative and formative evaluation was (and I believe is) especially useful in this arena. There is a substantial literature on normative evaluation and a much smaller and equally rigorous literature of the effects of formative feedback strategies. Our campus evaluation instrument is a direct descendant of the rigorous normative literature. However, to my knowledge it hasn't been validated in local circumstances in at least 25 years. The use of formative tools has been in practice for 25 years but in quite informal settings. There is hardly any research or evaluation examining the efficacy of these tools locally.....although anecdotal data are quite positive. I have heard criticisms of the normative instrument for years usually by individuals who have low scores. While the validity research nationally would argue for continued use. The just criticism of the instrument is: there isn't very much variance at the top of the scale. So it is very difficult to differentiate between average and good performers. I believe the instrument is still quite useful in identifying extremely poor and extremely good performers.

2. From my experience on the campus promotion committee I would assert the evaluation of teaching is quite sloppy and unprofessional campus wide. There is an over reliance on the campus instrument and it is often misused and misinterpreted. The committee has no provision for including teaching evaluation expertise in the committee process. When I served there were several instances of gross misinterpretation of teaching dossiers. I view this situation as an opportunity for the School to lead with better practice.

3. The best document I ever read on evaluation of teaching was written by Ken Gros Louis on the development of a teaching portfolio. (Karen Hawlett or Joan Middendorf could probably locate it.) His 1970s- 80s argument anticipated recent evaluation literature on the value of portfolios in making judgments about complex behavior over an extended period of time.

4. As a department chair in the school I was shocked and disappointed by the small percentage of faculty who deliberately falsify teaching evaluation data. I became aware of direct manipulation of the forms, intimidation of students by standing near students filling out the form etc. In one case several years of falsification added up to a positive recommendation for promotion.

5. The proper use of portfolio data would require systematic faculty and administrative development efforts to insure even minimal reliability and validity. This obvious choice for improving evaluation hasn't been

widely used because reasonable implementation requires far more sophistication and resource allocation than is readily available. The over use of campus instrument is an unholy compromise driven by available resources for creating useful data.

6. Peer evaluation is notoriously biased in my experience. Friends or mentors on the one hand or advocates for denying tenure typically provide this data now. The process should be far more disciplined than occurs in contemporary practice.

Data from Departmental Focus Group Survey

Tiffany Roman and Lenny Sanchez

- 1) Consistency with statement:

| | |
|---|----------------------|
| True for tenure and promotion (Criterion based) | More Prominent theme |
| Multiple forms of assessment are not necessarily used for annual review | More Prominent theme |
| There are many problems/confusions with current system. Faculty had many questions. | More Prominent theme |

- 2) How statement differs:

| | |
|---|----------------------|
| Merit review is not consistent with the statement (philosophically yes, but less materials permitted) | More Prominent theme |
| Annual reviews do not specifically require peer reviews | More Prominent theme |
| Considerations needed for online or blended courses, online documentation | Less Prominent theme |
| Disproportionate emphasis on student evaluations | More Prominent theme |
| Look at long term (5 to 10 year) impact on student ratings | Less Prominent theme |
| Student ratings are not global (2.5 & 3.5 are different, but 3.2 & 3.5 are rather similar) | Less Prominent theme |
| Student learning not emphasized to the degree it could be (only two items on BEST) | Less Prominent theme |
| Syllabi are not regularly turned in for review | Less Prominent theme |
| Much evidence relies on initiative of instructor (not required by department or school) | More Prominent theme |

- 3) To what extent are you satisfied with the approach to evaluating teaching in your department? Please comment about the extent of your satisfaction with respect to each of these areas: (1) merit review for faculty; (2) tenure and promotion of faculty; and, (3) hiring and retention of associate instructors

Merit review for faculty:

| | |
|--|----------------------|
| Faculty not satisfied with current system – more so with merit review than T & P | More Prominent theme |
| Confusions on what to provide for review | More Prominent theme |
| Most concerns are for junior faculty | More Prominent theme |
| Cutback on peer reviews due to time and concerns of valuable/reliable materials | Less Prominent theme |
| Unsure of how much weight is given to each category | More Prominent theme |

| | |
|--|----------------------|
| (e.g. student evaluations, syllabi, unsolicited notes and emails). | |
| Evidence needs to be more output than input oriented (think NCATE), though evidence may take years | Less Prominent theme |

Promotions and tenure for faculty:

| | |
|---|----------------------|
| Research is primary focus, teaching is valued far less | Less Prominent theme |
| Promotion and Tenure is criteria based/Merit review is norm based. | Less Prominent theme |
| Articulate differences between annual review criteria and P&T criteria (make process transparent) | More Prominent theme |
| P&T process is separate from annual review, so committee doesn't see annual or 4 th year review | Less Prominent theme |
| Promotion and Tenure portfolio is very time consuming. | Less Prominent theme |
| It's consistent and satisfactory as it genuinely requires multiple types of evidence, but annual review process is not excellent | More Prominent theme |
| Origin of source of evidence (school, department, individual faculty member, etc) may affect how promotion and tenure faculties compare across the SoE. | Less Prominent theme |
| Consider Gender bias, rank bias, subject matter – advanced vs. introductory | Less Prominent theme |

Hiring and retention of associate instructors:

| | |
|---|----------------------|
| Very few concerns with AIs | More Prominent theme |
| AIs may be teaching for 1 st time and succumb to student pressure. Tougher graders are penalized if others are grading easier. | Less Prominent theme |
| BEST scores seem to be heavily weighted. Low scores are a red flag and can haunt someone teaching a course for the 1 st time. | More Prominent theme |
| It's a learning role; our assessment has never had much bearing on their retention. Welcome efforts to systemize their assessment | Less Prominent theme |
| Look at mid-term evaluations for AIs (hard to evaluate AIs when teaching for the 1 st time) | Less Prominent theme |
| Need to prepare graduate students and AIs how to teach (as it's not done) | Less Prominent theme |

- 4) What changes, if any, would you like to see in the way teaching is currently being evaluated in your department?

| | |
|--|----------------------|
| There are problems, but it is unclear or difficult to solve them. | More Prominent theme |
| Certain courses tend to elicit lower evaluations (consider grouping courses into one category and evaluating as a group) | Less Prominent theme |
| Good teaching may be considered more seriously depending upon the department (hard to compare to publications) | Less Prominent theme |
| Mentoring of doctoral students should also be taken into consideration | Less Prominent theme |
| Specify where evidence should be located in the review (general specifications at moment) | Less Prominent theme |
| Merit review should foster culture of continual improvement, critique, and support | More Prominent theme |
| Consider comparing evaluations to evaluations conducted at other universities | Less Prominent theme |
| Consider summative observations by neutral expert colleagues (peer review) trained to evaluate teaching, but need a systemized process | More Prominent theme |
| Required to have five specific items on the course evaluations | Less Prominent theme |
| Consider range of scores for course evaluations (e.g. >3 o.k. <2.5 is an issue) | Less Prominent theme |

5) Perceived obstacles in revising the way teaching is currently evaluated in your department:

| | |
|---|----------------------|
| Clinical faculty practices are different from other professors, which impacts evaluations | Less Prominent theme |
| Reviewers need more guidance (know what to examine or where the evidence is located, additional criteria, and fine distinctions). | Less Prominent theme |
| Only a certain percentage can get an excellent rating – so we already start out with a limitation | Less Prominent theme |
| Different kinds of teaching reflect the need for standardization of the materials that should be included and submitted (e.g. do you teach this course every semester, kinds of students in the course, etc). | Less Prominent theme |
| Distant Ed courses and evaluations (not all evaluations turned in, potential biases) | Less Prominent theme |
| Issues of ‘Inherited practices’ (e.g. syllabi and routines that have been established over time and new person cannot be evaluated on the same practices) | Less Prominent theme |
| Evaluations for independent studies and advising -- these aren’t generally figured in but they can constitute a large proportion of teaching load. | Less Prominent theme |
| Time required with peer review is an obstacle | Less Prominent theme |
| To address teaching and improving the areas that are problematic, you need repeated measures (which takes several years) | Less Prominent theme |

6) How obstacles can be overcome and resolved:

| | |
|--|----------------------|
| Reduce requirements to three to four things | Less Prominent theme |
| Consider formative/summative outside or peer evaluation | Less Prominent theme |
| Make everyone aware that instructors can choose which BEST items to use | Less Prominent theme |
| Provide more details in explanations given in review findings | Less Prominent theme |
| First year evaluations should not be weighted so heavily | Less Prominent theme |
| Create rubric for merit review (consider requirements for Trustees Award with statement) | Less Prominent theme |

7) Additional comments:

| | |
|--|----------------------|
| The goal of evaluations is: increased pay or feedback to improve teaching? | Less Prominent theme |
| Specify how to produce outcome evidence and the teaching center within the school is primarily focused on technology | Less Prominent theme |
| Campus resources are more helpful but the school itself does not have anything | Less Prominent theme |
| First year evaluations should not be weighted so heavily | Less Prominent theme |
| Create rubric for merit review (consider requirements for Trustees Award w/statement) | Less Prominent theme |
| Whatever changes are made, let everyone know as soon as possible so preparations can be made for next year's review | Less Prominent theme |

Data from Departmental Chairperson Interviews

Tiffany Roman and Lenny Sanchez

1) Consistency with statement:

| | |
|---|----------------------|
| Evaluation differs between Tenure & Promotion and Annual Review | More Prominent theme |
| Consistency seems true for tenure and promotion (Criterion based) | More Prominent theme |
| Standardization does not occur during annual review process | Less Prominent theme |
| Purpose of Annual review is different: <ul style="list-style-type: none"> o Norm based o BEST scores, teaching publications, awards | More Prominent theme |

2) How teaching is evaluated in your department:

| | |
|---|----------------------|
| Multiple kinds of assessment are collected, but more so for T &P | More Prominent theme |
| BEST/ Course Evaluations | More Prominent theme |
| Teaching Articles | More Prominent theme |
| Collective faculty review of course content | Less Prominent theme |
| Course Syllabus | Less Prominent theme |
| List of participation on dissertation, thesis and program of study committees | Less Prominent theme |
| Teaching related activities (e.g., guest lectures, writing recommendations) | Less Prominent theme |
| Awards | Less Prominent theme |

3) Satisfaction with the approach to evaluating teaching described:

Merit review for faculty:

| | |
|---|----------------------|
| There are definite problems with current system, more so with merit review. | More Prominent theme |
| Streamline info for annual review (concise with multiple sources) | Less Prominent theme |
| Course evaluations are problematic (multiple reasons given) | Less Prominent theme |
| Articulate differences between annual review criteria and P&T criteria (make process transparent) | Less Prominent theme |

Promotions and tenure for faculty:

| | |
|--|----------------------|
| Good as is | Less Prominent theme |
| Course evaluations are problematic (multiple reasons | Less Prominent theme |

| | |
|---|----------------------|
| given) | |
| Articulate differences between annual review criteria and P&T criteria (make process transparent) | Less Prominent theme |

Hiring and retention of associate instructors:

| | |
|--|----------------------|
| Should require teaching evaluations (post in job announcement) | Less Prominent theme |
|--|----------------------|

4) Perceived obstacles in revising the way teaching is currently being evaluated in your department:

| | |
|---|----------------------|
| There are definite obstacles. | More Prominent theme |
| Different views/concerns how potential changes at departmental level could conflict with policies at school or university level | Less Prominent theme |
| Time and energy is required to change current systems and expectations (immense effort required for meaningful change) | Less Prominent theme |
| Requires online evals for DE classes so that instructors can get more feedback | Less Prominent theme |

5) How obstacles can be overcome:

| | |
|---|----------------------|
| Different views/concerns how potential changes at departmental level could conflict with policies at school or university level | Less Prominent theme |
| Time and energy is required to change current systems and expectations (immense effort required for meaningful change) | Less Prominent theme |
| Requires online evals for DE classes so that instructors can get more feedback | Less Prominent theme |

Data from Student Focus Group (Dean's Advisory Council) Survey

Wendy Wall-Marencik and Rajat Chadha

February 4, 2009

Demographics of Student Advisory Board Participants

| Total N=16 | Elementary Education | Teaching All Learners | Secondary Education |
|---------------|----------------------|-----------------------|---------------------|
| Seniors N=8 | 4 | 3 | 0 |
| Juniors N=6 | 2 | 3 | 2 (Eng/Math) |
| Sophomore N=1 | 1 | 0 | 0 |

(Missing=1)

1. Does the BEST course evaluation instrument (the standard computer-scanned survey), as it exists, adequately evaluate the teaching effectiveness of faculty and associate instructors?
2. Does the BEST instrument give you an adequate opportunity to express your learning? Please explain.

| Total N=16 | Elementary Education | Teaching All Learners | Secondary Education | Unknown |
|------------|----------------------|-----------------------|---------------------|---------|
| No | 7 | 5 | 2 | 1 |
| Yes | 0 | 1 | 0 | 0 |

Summary and Themes:

All but one of the participants (94%), believe the BEST does NOT adequately evaluate teaching effectiveness of faculty and associate instructors or express their learning. Reasons cited for this overwhelming agreement were predominately the non-specificity of the items and not personalized to the instructor (e.g., "No, because I think the questions asked do not always apply to the specific teacher or course", "Assessment is very general"). Another common response regarded the "timing" in which the BEST was given, (e.g., "At the end of the period on the last day, no one wants to fill this "extra" stuff out", "I feel like I'm rushed to complete right after exams"). To a lesser degree some concern was expressed about teacher measurement versus student learning.

3. Does the BEST instrument give you an adequate opportunity to express your estimation of an instructor's teaching? Please explain.

| Total N=16 | Elementary Education | Teaching All Learners | Secondary Education |
|------------|----------------------|-----------------------|---------------------|
| No | 5 | 4 | 1 |
| Yes | 0 | 2 | 0 |
| Too vague* | 2 | 0 | 1 |

*Not specifically a yes or no

Summary and Themes:

The majority of participants selected “No” and again cited “lack of specific detail about the instructor or course” in the survey questions and “limited general questions with only five options for rating” as problematic in having an adequate opportunity to express their estimation of an instructor’s teaching. There appeared a general desire to have more options for providing narrative feedback, (e.g., “Wish we had more space to explain our choices during the survey. Maybe a question of how they can improve?”) And again, a few comments about not having enough space or time to complete an adequate write-up.

4. If not, what alternate means of evaluating would you suggest?

Summary and themes:

The most frequent response to this question was consideration for online availability throughout the semester or at various times. According to various respondents, this would assist with the issue of time in completing the evaluations. Others less frequent suggestions were consideration for a short answer format, more open-ended questions, essay or free response type format and more space to write.

5. Would your peers agree that the BEST course evaluation instrument, as it exists, provides them an opportunity to express their estimation of their learning? Please explain.

6. Would your peers agree that the BEST course evaluation instrument, as it exists, provides them an opportunity to express their estimation of their instructor’s teaching? Please explain.

| Total N=16 | Elementary Education | Teaching All Learners | Secondary Education |
|------------|----------------------|-----------------------|---------------------|
| No | 5 | 3 | 1 |
| Yes | | 1 | |
| Too vague | | 2 | 1 |

Summary and Themes:

The explanations for question 5 and 6 seemed to be together. The majority of students responded “no” that they did not believe their peers would agree the BEST provides them an opportunity to express their estimation of their learning or their instructors teaching. No consistent themes were indicated, however several individuals responded that students rushed through these survey and did not take time to truly evaluate their learning or teacher’s instruction (e.g., “Most students rush through their evaluations and don’t know how detailed to be.”). Another response by several individuals related to providing better questions and a more narrative response format, (e.g., “No, I feel that several of my peers would prefer to explain how their teacher did over the semester as opposed to ranking it on a scale”, “When teachers add more specific questions it is helpful”, No, the assessment is always given at the end of the class and students want to get out and never truly fill out the assessment and they don’t like the lack of open ended questions).

(Did not see data on this)