

A COMPARISON OF FACULTY AND STUDENT PERSPECTIVES ON COURSE EVALUATION TERMINOLOGY

Carol Lauer
Rollins College

Students completed a questionnaire on the meanings of several terms commonly used in narrative course evaluation forms, and instructors completed a similar form asking them what they believed students meant by these same words. Both groups produced complex and multifaceted definitions for the terms "not fair," "professional," "not organized," "challenging," and "not respectful." Instructors were frequently incorrect in inferring common student definitions. If instructors misjudge students' meanings, how useful are evaluation forms for evaluative or developmental purposes? Faculty peer review of class meetings and course materials can help provide the necessary filter for interpreting evaluation comments.

○

I once believed that narrative comments on course evaluation forms were straightforward and useful. Then I had the opportunity, as a department chair and as a member of a collegewide tenure and evaluation committee, to read literally thousands of course evaluations for faculty at all stages of their careers.

At the same time, I examined syllabi and other course materials for the faculty under review and attended a sampling of their classes. I quickly discovered that narrative comments on student evaluations forms were far from straightforward and could not be easily connected with course materials and my class observations. Because student course evaluations are frequently used for evaluative and developmental purposes, this would appear unwise without some decoding process if they are as cryptic as my experience suggests.

While a great deal of research has been done on the problems of the numerical rating systems in course evaluations (Benton & Cashin, 2012; Gray & Bergmann, 2003; Laube, Massoni, Sprague, & Ferber, 2007; Pounder, 2007), a dearth of information exists on how to interpret student comments. This study examined what students at a single institution meant when they used several commonly seen terms in narrative sections of evaluation forms and what faculty believed students meant by these terms.

The Literature on Limitations of Student Course Evaluations

Most institutions use evaluations with numbered scales (Laube et al., 2007) since the ratings are easily summarized and compared. This practice continues to grow despite studies that show ratings can be biased by a variety of factors that have nothing to do with good classroom practices and resultant student achievement (Gray & Bergmann, 2003; Pounder, 2007).

Pounder (2007) summarized a wide range of studies demonstrating that class size, age, experience, and gender of instructors have an impact on numeric scores on course evaluations. Since instructors have no control over any of these factors, they provide few useful insights for developmental or evaluative purposes. Evaluations may even punish instructors for policies that other faculty would view as appropriate and even exemplary. For example, instructors whom students view as giving too much work and grading harshly are rated lower than their less demanding peers (Laube et al., 2007). The literature also documents a host of seemingly random effects that can influence evaluation results. When expected grades were low, students gave lower ratings to female instructors than to male instructors (Laube et al., 2007). Faculty who bring food to class are rated higher than those who do not (Pounder 2007). Better-looking instructors get higher ratings (Gray & Bergmann, 2003). Faculty in the arts and humanities are rated as more effective teachers than those in more quantitative fields like economics, math, and physics (Pounder, 2007).

Other findings bring into question the seriousness with which students fill out their evaluations. Beyers (2008) argued that students pick a number they feel represents their instructor's level of achievement and use it repeatedly regardless of the question posed. Beyers based this conclusion on observations in large lecture halls packed with students filling out course evaluation forms. He also conducted a personal experiment where he returned every class assignment at the class meeting immediately following its submission, and yet only five of his thirty-nine students give him the highest numerical rating for "returned materials promptly." Rating instructors is clearly not a straightforward enterprise. Benton and Cashin's (2012) summary of forty years of research on student ratings is also revealing in the number of disagreements and alternative hypotheses presented. They suggest the use of "multiple sources of information" and the careful interpretation of results (p. 13).

Although these studies focus on numbered scale evaluations, my own reading of course evaluation forms suggests that similar biases affect narrative comments. For example, arts and humanities professors are more likely to receive superlative comments on their forms than are instructors in other disciplines, and the descriptors used for female faculty are likely to be different and, under some circumstances, more negative than those used for males. While the existing literature can help faculty and their evaluators deconstruct student comments in the light of these likely biases, what about the issue of the terms used to describe instructors? Can we assume they have a set of standard meanings that students and instructors share? Could faculty be misinterpreting these terms?

The Study

Methods

My institution uses a narrative evaluation form as well as a numerical ratings scheme for student course evaluations. The narrative form includes open-ended questions like: "What, if anything, could have been changed about this course to improve your learning experience?" "What were the major strengths and weaknesses of the instructor?" An initial review of the comments on forms for several hundred classes and close to seventy instructors, along with reviews of syllabi, other course materials, and classroom visits, suggested that some commonly used terms must have more than one meaning. An instructor with a syllabus that includes few specifics and a nonstructured approach to classroom discussions might be labeled

as disorganized, but so might an instructor with a clear and detailed syllabus and a highly structured classroom. An instructor who uses a detailed rubric for grading might be called unfair in grading practices.

The study presented here examined whether a set of commonly used terms means different things to different students and whether a single term could have multiple meanings for a student. A second line of inquiry was whether faculty accurately predicted student meanings for this set of terms and if they saw these terms as having single or multiple meanings. Finally examined are faculty and student agreement between and among them about whether the attributes described are important.

Research Design

A set of five common terms was culled from narrative comments on student course evaluations for use in a survey of students and faculty at a medium-sized liberal arts college: *not fair*, *professional*, *not organized*, *challenging*, and *not respectful*. Students used these terms frequently, yet interpreting what they meant continually challenged the collegewide tenure and promotion committee. Questionnaires about the meanings of the terms were developed for a nonrandom sample of students and faculty and were approved by the college's institutional review board. The questionnaires are reproduced in Exhibits 13.1 and 13.2. Both negative and positive terms were presented to avoid the appearance of bias toward a particular view of evaluation comments.

Exhibit 13.1. Student Questionnaire.

Year: Please circle one Freshman Sophomore Junior Senior
Gender: Please circle one Male Female
Major:
What do the following phrases from course evaluation forms mean to you? The instructor is: Not fair Professional Not organized Challenging Not respectful
Are any of these critical to the success of a course? If so, which ones?

13.2. Faculty Questionnaire.

Rank: Please circle one Instructor Assistant Professor Associate Professor Full Professor
Gender: Please circle one Male Female
Division: Please circle one Arts Humanities Social Sciences Science and Mathematics
What do the following phrases from course evaluation forms mean to you? That is, what do you think students are trying to convey when they use these phrases. The instructor is: Not fair Professional Not organized Challenging Not respectful
Are any of these critical to the success of a course? If so, which ones?

Student questionnaires were handed out during spring semester so that all students had had some experience filling out the college's standard course evaluation forms during the fall. Seven instructors representing a variety of disciplines and all four undergraduate divisions (Arts, Humanities, Science and Mathematics, and Social Sciences) handed out and collected the anonymous forms during their own class sessions. The courses involved from first year to senior level. Faculty questionnaires, which were also anonymous, were passed out during several faculty meetings and collected by the end of the meetings. These meetings included faculty representing all undergraduate divisions and departments.

Of the 185 students who completed the questionnaire, 68 percent were female and 32 percent were male; 32 percent of them were first-year students, 20 percent sophomores, 20 percent juniors, 23 percent seniors, and the rest were of unknown status. They were majors in every undergraduate division.

The 45 faculty members who responded were 53 percent female and 44 percent male, with the rest not reporting gender. They represented every rank, with 9 percent of them holding the rank of instructor, 35 percent assistant professor, 27 percent associate professor, and 29 percent full professor. The rest did not indicate rank. Respondents came from all four undergraduate divisions.

Data Analysis

I did all coding. The open-ended portions of the questionnaires were coded for key word phrases (Bernard, 2011). The coding process at this level was frequently straightforward since many respondents used the same terminology. A single request for a definition on the questionnaire frequently produced multiple phrases in response, and each was coded independently so that one answer might generate several entries for the coding book. Once all the questionnaires had been coded, key word phrases were collapsed into broader thematic categories using an open coding system. That is, the themes emerged from the data (Bernard, 2011).

To preserve the diversity of answers, thematic categories were kept fairly narrow. For example, “plays favorites,” one of the themes listed for the term *unfair*, was mentioned in 59 of the 119 student responses in this category. The variant phrase “unequal treatment of students” appeared 15 times, and “biased toward certain students” appeared 7 times. Since all these phrases concerned differential treatment of students, they were considered part of a single theme. However, not all coding was this simple. The theme “formal” that emerged for the term *professional* included “serious,” “businesslike,” “a balance between friendly and authoritarian,” and “designated adult.” Although no single phrase dominated in the responses, all definitions focused on a particular way of behaving in a classroom.

Results

As Table 13.1 demonstrates, students and faculty listed multiple definitions for the terms presented. Both groups generated the most multifaceted definitions for the terms *professional*, *not organized*, and *not respectful* despite the fact that *organized* and *respectful* are defined on the numerical ratings portion of the standard course evaluation forms used by every class. Examining the diversity of definitions for each term provided insight into agreements and misunderstandings within and between faculty and student groups.

Table 13.2 presents results for the term *not fair*. In Tables 13.2 through 13.7 the heading “number” refers to the total number of times a particular definition category or theme was listed, and “percentage” converts this number into a percentage of the total number of definitions listed. For all tables, only categories that included at least 5 percent of the key word phrases recorded are listed. For the term *not fair*, most student responses fit into four themes. Almost half of the students used the term to mean that the professor has favorites in the class whom he or she treats in a

Table 13.1. Survey Results.

Term	Students			Faculty		
	Number of Responses	Number of Respondents	Average Number of Responses per Individual	Number of Responses	Number of Respondents	Average Number of Responses per Individual
Not fair	260	177	1.5	63	44	1.4
Professional	360	174	2.1	81	43	1.9
Not organized	353	176	2	80	39	2.1
Challenging	255	177	1.4	68	44	1.5
Not respectful	312	167	1.9	78	39	2.0

Table 13.2. What *Not Fair* Means to Students.

Comments	Students		Faculty	
	Number	Percentage	Number	Percentage
Plays favorites	119	45.8	20	31.7
Grading problematic	58	2.3	31	49.2
Work is too hard	33	12.7	0	0
Will not "work with you" on problems	32	12.3	0	0
Other	18	6.9	12	19.0
Total	260	100	63	100

manner that privileges them. Another 35 percent saw the term as meaning a course that is overly difficult, which could refer to harsh grading, unclear grading, or demanding course content. These two themes are broken out in Table 13.2. The other prominent category focused on a professor's willingness to bend the rules on request. This could mean letting a student redo an assignment, allowing extra time for assignments, or allowing a student to skip an assignment when family concerns or other work interfered—for example, "Does not try to work around any problems you may have with an assignment or something." The irony is that instructors are "not fair" if they do or do not make exceptions for students. Faculty thought students usually used the term to mean one of two things: grading problematically (usually too harsh) or favoritism. No one listed hard work or an unwillingness to bend the rules to accommodate a particular student's needs as a possible student definition.

Results for the term *professional* appear in Table 13.3. Much less agreement than for *not fair* existed on what this term meant for both faculty and student groups. No definition rose above the 10 percent mark for students, and ten terms had a frequency of over 5 percent. The definitions listed focus on a wide range of items, from being knowledgeable and formal in class to being on time and wearing, literally, a suit. The formal comportment category, for example, includes "uptight," "air of dignity," "knows the difference between inside and outside of class behavior," and "remains the designated adult in class." Ironically, given the lack of consensus on meanings that Table 13.3 demonstrates, this is one of the attributes that students found most critical for the success of a class, as will be discussed in reference to Table 13.7.

Table 13.3. What *Professional* Means to Students.

Comments	Students		Faculty	
	Number	Percentage	Number	Percentage
Knowledgeable	37	10.3	7	8.6
Formal comportment	35	9.7	11	13.6
Timely and efficient	31	8.6	10	12.3
Organized	29	8.1	6	7.4
Calm and polite (includes language)	29	8.1	5	6.2
Fair and unbiased	28	7.8	2	2.5
Prepared	26	7.2	5	6.2
Not too friendly with students	24	6.7	5	6.2
Controls classroom	23	6.4	0	0
Respects students	22	6.1	4	4.9
Dresses formally (suit)	16	4.4	10	12.3
Other	60	16.6	16	19.8
Total	360	100	81	100

For faculty, two definitions, "formal comportment" and "timely and efficient," rose above the 10 percent level, and eight had frequencies higher than 5 percent. Faculty, like students, presented many definitions for the term *professional*, and they often appear at frequencies similar to those found in the data from students. The primary exception is that instructors believed students were much more focused on how faculty members dress than students actually reported they were.

This term has many responses (seventy-six) coded as "other," which illustrates the breadth of answers that may make up this category. Ten students said it meant "helpful." Another six defined it as "available." Singular student responses included "updates Blackboard with grades," "lectures," and, "unattached." Faculty responses included "doesn't forget things," "strict," and "traditional."

Table 13.4 provides data for *not organized*. Students' responses produced nine categories that included 5 percent or more of the definitions, but faculty responses fit mostly into only three of these popular categories. For students, the most common definitions were that faculty members are not prepared, do not have a clear plan for the class period, or are slow to return graded tests or assignments. Many comments, however, concerned changing syllabi, losing papers, and forgetting things like assignment due dates.

Table 13.4. What *Not Organized* Means to Students.

Comments	Students		Faculty	
	Number	Percentage	Number	Percentage
Not prepared	61	17.3	10	12.5
No apparent daily plan	53	15.0	4	5.0
Grades slowly	45	12.7	4	5.0
Changes or does not follow syllabus	40	11.3	23	28.8
Forgetful	31	8.8	2	2.5
Loses things	29	8.2	2	2.5
Late	28	7.9	4	5.0
Unclear syllabi, assignments	25	7.1	5	6.3
Goes off topic	1	—	7	8.8
Other	41	11.6	19	23.8
Total	353	100	80	100

Only two faculty members, however, mentioned knowledge of due dates on their questionnaires. For faculty respondents, who had only four categories above the 5 percent mark, close to 30 percent of responses were about changing or not following the syllabus. They believed that students used this term to describe someone who gets off topic in class, but only 1 of 353 student phrases about organization focused on this concern. Only four faculty (5 percent of responses) mentioned daily plans as associated with organization in student minds (as opposed to 15 percent for students), and slow grading appeared only forty times (5 percent) in eighty key word phrases versus representing 12.7 percent of student responses.

While the "other" category for this term is not as large (sixty responses) as that for *professional*, the variety of responses demonstrates how idiosyncratic answers could be. Six students responded with some version of "is messy": "disheveled," "disorderly office," and "disorderly course materials." "Ditsy with PowerPoints" is an example of a particularly opaque comment. Faculty members also recorded singular responses: "a normal human," "not obsessive," "shy," and "self-deprecating."

Students and faculty had much more unified views of the term *challenging*, as is evident in Table 13.5. For students, six categories rose above the 5 percent mark, but four of these could be collapsed into the broad theme of "hard work," which constitutes 57.3 percent of all the

Table 13.5. What *Challenging* Means to Students.

Comments	Students		Faculty	
	Number	Percentage	Number	Percentage
Thinking required	4	21.2	5	7.3
Hard work combined	146	57.3	31	45.6
Hard work	51	20.0	18	26.5
Lots of work	39	13.5	8	11.8
Hard but reasonable	31	12.2	5	7.3
Too hard	25	9.8	0	0
Hard grader	18	7.1	17	25.0
Engaging	15	5.9	3	4.4
Other	22	8.6	12	17.6
Total	255	100	68	100

responses. I separated out the subcategories since over 20 percent of these were surprising, to me at least, in that students felt the need to qualify the kind of hard work to be "hard but not too hard" (12.2 percent) or to say that challenging was not a good thing (9.8 percent) since it meant that the work was too hard.

For faculty, five categories were above the 5 percent mark, but 45.6 percent could be placed in the "hard work" category. Faculty members, unlike students, never understood "challenging" to be a negative attribute meaning "too hard." Faculty were also almost four times more likely than students to include a reference to hard grading in their list of definitions. They were three times less likely than students to make a reference to critical thinking in their list. This may mean that many faculty are unaware of the importance of critical thinking to students.

Finally, the results for the term *not respectful* are shown in Table 13.6. Student responses are clustered in five categories and faculty in three. Students' top category focused on instructors not listening to them or taking them seriously. The comment "treats us like children" or "like an inferior species" came up repeatedly (seven times). "Criticizes student ideas" was also part of this set of responses and constituted thirteen of the ninety-eight comments included in this grouping. The next most frequently listed definitions were "embarrasses students" or "demeans students" and was "rude to students." These categories are listed separately but may well be all about the same set of behaviors. Since the word *rude*

Table 13.6. What *Not Respectful* Means to Students.

Comments	Students		Faculty	
	Number	Percentage	Number	Percentage
Does not listen to or does not take students seriously	91	31.4	23	29.5
Embarrasses or demeans students	49	15.7	18	23.1
Rude to students	48	15.4	8	10.3
Not understanding of student problems	34	10.9	5	6.4
Discriminates against certain students	24	7.7	1	1.3
Other	49	15.7	23	29.5
Total	312	100	78	100

was usually listed with no qualifiers, it was hard to do more than guess as to its meaning. "Not being understanding of student problems" came up here (10.9 percent) and also as a definition for the term "fair."

The top three categories for faculty members' sense of what students meant by "not respectful" mirrored those of the students. However, 14 percent (eleven) of faculty responses mentioned criticizing student ideas as a definition for this term, making this definition more than three times more frequent than in the student responses. For faculty 29.5 percent of all responses fell into the "other" category, meaning they could not be easily grouped. Included were comments like "blunt," "does not call students by name," and "does not act maternal," which are not mirrored on the student list.

The final question posed to both students and faculty was whether any of these terms were about attributes critical to the success of a course. Few students (five) or faculty (three) thought the terms were unimportant, while fifty-three (28.5 percent) students and eleven (24.4 percent) faculty thought all were important. As Table 13.7 shows, for students, the most critical categories were "professional," "not organized," and "challenging." "Professional" and "not organized" also had the most definitions listed by students. Combining votes for the individual category with those for the "all" category generated these percentages. For example, the 118 student votes for "professional" include 65 separate votes for "professional" and 53 votes for the "all" category, which, of course, includes "professional."

Table 13.7. Attributes Considered Critical for Course Success.

Comments	Students		Faculty	
	Number	Percentage	Number	Percentage
Professional	118	21.00	17	14.8
Organized	116	20.6	24	20.9
Challenging	114	20.3	27	23.5
Fair	106	8.9	16	13.9
Respectful	95	16.9	19	16.5
None of these	5	0.9	3	2.6
Left blank	8	1.4	9	7.8
Total	562	100	115	100

Faculty thought the most critical attributes for course success had to do with organizing and challenging students. They did not follow the students' pattern of more definitions for the terms they felt most important. They did agree with students on "organized" and "challenging" as the most important of the attributes under consideration.

Discussion

The results of this study show that faculty and students do not have synonymous definitions of a set of terms that are common on narrative course evaluation forms. For all the terms, multiple definitions, whose meanings may or may not overlap, were generated by both faculty and student samples. Faculty members believed students had many definitions for these terms (and students did); faculty sometimes accurately identified student definitions but frequently did not. In actuality, faculty and students definitions were frequently at odds and can be quite idiosyncratic. It might be that while no consensus on meaning emerged, it is because these are not considered important terms. It could even be that their lack of significance led to the multiplicity of definitions. That is, they were characteristics that people had not given much thought to explaining.

Both groups, however, saw at least some of these terms as representing characteristics critical to course success. It is therefore particularly unfortunate that the terms are so opaque. As noted, two of the most critical categories for students were "professional" and "not organized," and yet these also had the most diverse definitions. Perhaps they are considered so important because they encompass so many attributes.

Instructors have the job of reading these forms and trying to figure out how to improve their teaching based on the comments—not easy to do since the meanings are so inconsistent. Many instructors thought that students labeled them as disorganized because they changed their syllabi over the course of a semester, when student comments were really about returning graded papers promptly. The problem can be even worse in evaluations based on numeric rating scales since the narrative comments sometimes include qualifying descriptions or examples that help faculty decipher meanings, something that rating scales never do. Clearly faculty in this small sample had not been well educated by such contextualizing student comments because they frequently misunderstood what students meant by the basic terms surveyed.

Beyond instructors, administrators and evaluation committees at the college under review use narrative descriptions for developmental and evaluative purposes. This study suggests how problematic that can be and has implications for the similar use of numeric scales. My own experience, and that of my colleagues on our collegewide evaluation committee, indicates a great deal of confusion in trying to match what we observed about instructors with what we read in comments. We eventually learned that without class visits, we could not attempt any assessment of an instructor's work as a teacher.

A single example may prove instructive. I read the evaluations of a candidate who was clearly liked and respected by students but was frequently labeled by them as disorganized. The individual had carefully designed Web sites for courses and clearly took great care with syllabi and other course materials. The instructor was plainly confused about what the comments about organization could mean. I attended a class meeting and quickly realized that while the professor was organized on the macrolevel, he was not on the microlevel. He brought in the wrong memory stick, for example, apologized profusely, and returned to an office to fetch the appropriate one. Since the office was only a few steps away, the class still started on time. The instructor mentioned forgetting things several times, but if not for the instructor's self-deprecating comments, I would never have noticed anything was amiss without having attended the class and, without this direct classroom experience, could not have helped the instructor understand the comments on evaluations. I learned to decipher student assessment of this instructor as meaning, "This is someone who does an excellent job, but has some organization issues, albeit minor ones."

In this case I came to agree with the students, but understanding what students meant by the terms surveyed in this study suggests that students

sometimes define excellence in teaching in ways with which most faculty could not agree. Some of the higher-frequency definitions of terms that students label as critical for course success are reasonable: be prepared, grade promptly, have students work hard (although not too hard) with some critical thinking involved. Other frequent comments, however, fit with a customer satisfaction model for evaluations but are not helpful for improving teaching (for example, "Willing to let things slide when we have time management problems" as a definition of *fairness*).

Even some definitions that may appear reasonable at first glance, like "has no apparent daily plan" for "not organized," can reflect lack of understanding of a discipline. The instructor could have an unapparent daily plan, as in philosophy classes I observed where the "plan" is for student commentary on the readings to shape the focus of the discussion, with many alternatives possible and equally enriching. An instructor who follows this approach might be labeled "not organized" but is actually "alternatively organized." Without some lexicon of what terms might mean to students and visits to classroom, evaluators are not likely to be able to make these distinctions. Narrative course evaluation questions would be much more useful to evaluators and instructors if they asked for specific examples and not just general statements. This might help direct teaching energies toward changing patterns that need to be changed and not toward changing what are already good practices that simply need to be explained to students.

Finally, faculty and students did not entirely agree on what is critical for course success. More students than faculty, for example, thought that being "professional" and "fair" were important. If, however, faculty and students do not agree between or among themselves as to what these terms mean, how can instructors or evaluators use comments about them for any useful purpose? While almost everyone in the sample thought at least two of the terms (or all of them) surveyed represented something critical to course success, no consensus emerged about what is absolutely critical. For students the highest numbers were for "professional" and "organized" at 21 percent of responses and for faculty "challenging" with 23.5 percent of responses.

Many students' comments about the positive attributes of instructors fit into the theme of "student centered" abstracted by Onwuegbuzie et al. (2007) from a survey of over nine hundred students at a midsize public university. "Professional" also emerged as one of the metathemes uncovered by this study of student perceptions of effective teachers (Onwuegbuzie et al., 2007). While this suggests that the survey focused on some widely held student beliefs, unless faculty can decode these from

evaluation forms, they provide little help for instructors in setting priorities. In sum, without peer review of class meetings and course materials, and requests for clarifying information, especially examples, on course evaluation forms, comments are likely to be misinterpreted, and instructors are likely to spend time and energy trying to fix the wrong problems or nonexistent problems. Student comments, however, could be a rich source of information for faculty development if peer reviewers would educate faculty members about their complex nature.

Conclusion

The results of this study are clearly preliminary since they used a sample of students and faculty from a single institution. Yet while the idiosyncrasies of the definitions provided may represent campus culture, it would certainly be useful to check on the study's applicability to other campuses. Unless institutions do extensive training in the vocabulary of evaluations, it seems unlikely that the broad findings are unique to a single location.

These findings suggest that narrative comments on course evaluations can have many meanings to students and to faculty, and not necessarily the same meanings. Given this, their use without contextualizing information can lead to inaccurate evaluations and misplaced faculty efforts. To address this problem on the campus studied, a pilot project on an informed peer review process, organized by the campus's Institute for Effective Teaching, will include the discussion of the intended meanings of student course evaluation responses. Peer reviewers will attend classes of several faculty volunteers, examine course evaluations, and help faculty members interpret the meaning of student responses based on these classroom observations. While the bureaucracy required to change the evaluation of teaching procedures makes reform a slow process, we hope that this pilot project will facilitate such changes and lead to training faculty in deciphering intended student meanings.

REFERENCES

- Benton, S. L., & Cashin, W. E. (2012). *Student ratings of teaching: A summary of research and literature*. (IDEA Paper No. 50). Manhattan: Kansas State University, Center for Faculty Evaluation and Development.
- Bernard, H. R. (2011). *Research methods in anthropology* (5th ed.). Lanham, MD: Altamira Press.

- Beyers, C. (2008). The hermeneutics of student evaluations. *College Teaching*, 56(2), 102-106.
- Gray, M., & Bergmann, B. R. (2003). Student teaching evaluations: Inaccurate, demeaning, misused. *Academe*, 89(5), 44-46.
- Laube, H., Massoni, K., Sprague, J., & Ferber, A. L. (2007). The impact of gender on the evaluation of teaching: What we know and what we can do. *NWSA Journal*, 19(3), 87-104.
- Onwuegbuzie, A. J., Witcher, A. E., Collins K. M. T., Filer, J. D., Wiedmaier, C. D., & Moore, C. W. (2007). Students' perceptions of characteristics of effective college teachers: A validity study of a teaching evaluation form using a mixed-methods analysis. *American Educational Research Journal*, 44(1), 113-160.
- Pounder, J. S. (2007). Is student evaluation of teaching worthwhile? An analytical framework for answering the question. *Quality Assurance in Education*, 15(2), 178-191.