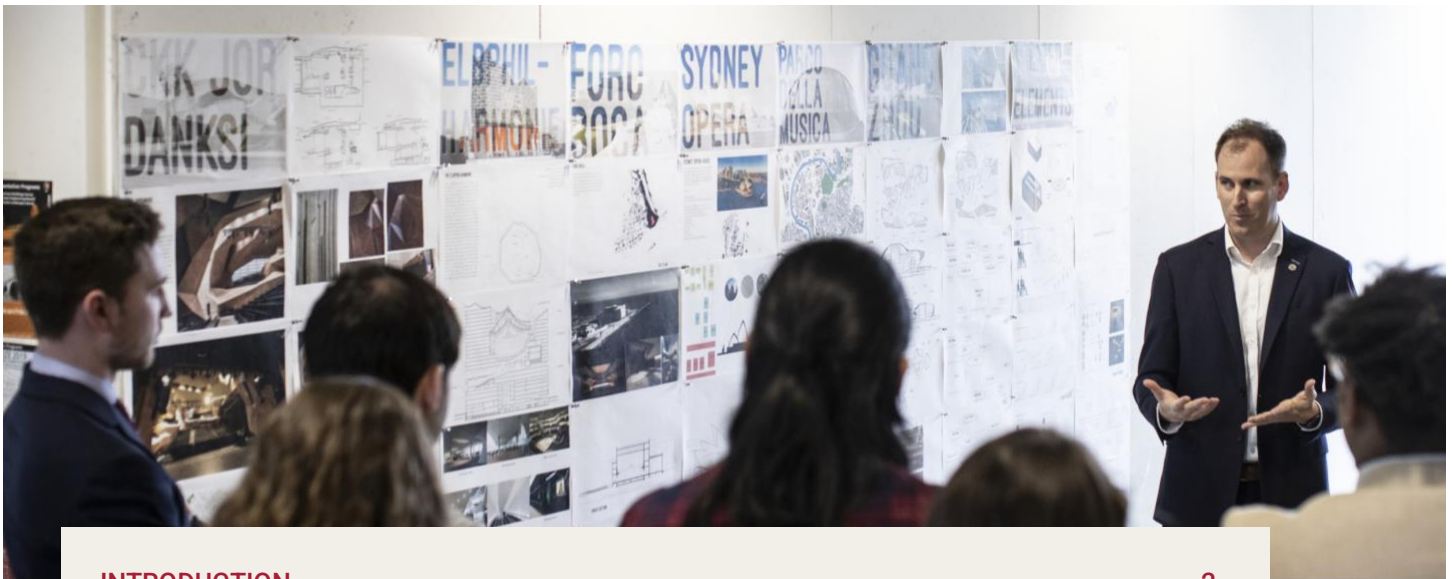


Recommendations for the use of Student Feedback Form (SFF) data at Temple University

Assessment of Instruction Committee Recommendations



INTRODUCTION	2
STUDENT EVALUATIONS OF TEACHING, AN OVERVIEW	3
WHAT STUDENT RATINGS ARE AND ARE NOT	4
ADVICE FOR INSTRUCTORS ON HOW TO MAKE THE BEST USE OF SFF DATA	6
GUIDELINES FOR THE USE OF SFF DATA TO EVALUATE INSTRUCTORS.....	8
ADVICE TO CHAIRS AND DEANS ON HOW TO SPEAK TO FACULTY ABOUT THEIR SFFS	11
CONCLUSION	13
REFERENCES	14

The Committee wishes to thank Dr. Angela Linse, Director of the Shreyer Institute for Teaching Excellence at Penn State University, and a national expert on student evaluations of teaching, who generously gave permission to use several of her publications in this document. Several of these are listed in the reference section and provide extensive reviews of scholarship in this area. A literature review, including Dr. Linse's research, is included for those who want to explore this area more deeply.

Introduction



In academic year 2015–2016, what was then the Student Feedback Form committee (now called the Assessment of Instruction Committee, AIC) approved the development of a new SFF form/system. After extensive review of the course and teaching evaluation practices at other institutions, an exploration of the latest research on student feedback forms, and a redesign of the SFF technology platform, this evaluation process culminated with changes that were piloted during the 2019–2020 academic year and were fully implemented during the fall 2020 semester. The changes included:

1. A change to the standard form with a reduced number of required evaluative items (now 4, down from 11).
2. The ability to customize the remainder of the form by selecting items from an item bank.
3. A change to administering the form via a newly developed web-based platform that allows flexible, streamlined administration and reporting.

The major enhancement from the existing online system has been to greatly increase the flexibility of the system by allowing items (questions) to be added at four levels: course attribute (General Education, Honors, Online, and Writing Intensive), college, department (subject code), and individual instructor. The AIC was asked to guide this process. As part of that enhancement, the AIC was also asked to create a document addressing the proper use of student feedback data. This document is the AIC's response to that request.

This document provides a set of guidelines and suggestions for the use of Student Feedback Form (SFF) data by individual faculty, faculty committees, school and college leadership, and university-level administrators at Temple. The document contains a brief history of the SFFs, a discussion of what these data represent, guidelines for their appropriate use, and suggestions for ways to get the most benefit from them. The document was created by the AIC as part of its charge to provide guidance on assessment to the University community. This is especially relevant in the current context where student ratings of instruction (commonly called Student Evaluations of Teaching or SETs in the literature) have come under increasing scrutiny and criticism.

Student evaluations of teaching, an overview



Student evaluations of teaching have been used in higher education for approximately 100 years. There are three basic reasons why SETs are used:

1. To improve teaching quality.
2. To provide input for appraisal exercises (e.g., tenure and promotion decisions, merit, contract renewals).
3. To collect evidence for institutional accountability.

These goals are at least sometimes in conflict with each other, and many criticisms of the use of SETs arise because of the competing nature of these goals. In education research, *formative* assessments or evaluation are those used for the development or improvement. By contrast, a *summative* assessment or evaluation involves making judgments about efficacy or quality at its conclusion. In the numerous criticisms of SETs, the issue is seldom the use of SETs as a tool for improving teacher or course quality. That is, the correct use of SETs by individual instructors, in a formative manner, to assess and then improve their teaching is a valuable method that all instructors should use, and this is supported by substantial research-based evidence. When SETs are used formatively, there is almost no criticism of their use. It is when SETs are used for summative purposes (i.e., to make evaluations, comparisons, or judgments), especially if they are used as the sole or major source of data, that problems arise. In summative contexts, quite a few issues have been raised about SETs (validity and reliability concerns, for example), but the major focus of the literature is on bias. On balance, the AIC recommends retaining SETs at Temple. In addition to their function in providing formative guidance to instructors, SETs also give students the ability to provide feedback to instructors, which is an important signal that the institution is concerned about the quality of its teaching and that it cares about what its students think.

This document is intended to help faculty interpret and use the data derived from Temple's SFF results to inform their own pedagogy and, more critically, when they are in evaluative roles such as serving on Merit or Promotion and Tenure Committees. As mentioned above, this document also provides guidance for administrators in their role as evaluators.

What student ratings are and are not



Before presenting recommendations about the best use of SETs, it is necessary to briefly discuss what these data represent. It must be stated at the onset that there is no topic in this area where there is complete consensus. Any of the statements made here can be contested. However, after more than 80 years of research on this topic, and after thousands of research articles have been published, there are some statements that can be made with more assurance than others. The following are those that the AIC believes meet this standard.

1. **Student ratings are student perception data.**

SETs represent the collective views of only that group of people who have experienced the class and who have chosen to report. This statement may seem obvious, but the word “only” in the previous sentence is important. These students comprise the only group that has had the opportunity to observe the instructor and how they have impacted the learning environment of the course. These are perceptions, but they are unique—other faculty cannot have the same depth of knowledge on which to make evaluative judgements. At the same time, they represent only the views of students who choose to report, which may be a problem depending on response rate.

2. **Student ratings are sometimes biased.**

The evidence for this assertion is clear: student ratings can be biased by characteristics such as race, gender, language competency, and attractiveness, among other factors. For example, some research shows that minority instructors receive lower ratings, and African American faculty receive the lowest evaluations of any other group. Likewise, studies have shown that women faculty often receive higher evaluations than male instructors. Additionally, instructors with pronounced accents or dialects also tend to score lower on student evaluations than mainstream US English-speakers.

3. **Student ratings do not measure student learning.**

After years of research on this topic, there is almost complete consensus that SETs cannot and should not be used as a proxy for student learning. This may appear counterintuitive because it would seem logical that students who learn more will provide higher ratings for both the course and the instructor. While that may be correct in certain cases, the point made above must be kept in mind: SETs are student perceptions, and perceptions can be and are based on many factors. Although one of these factors may be how much the student has learned, it should not be assumed that courses or instructors that obtain high SET ratings are necessarily those where the student has learned more, nor is the opposite necessarily true. This will be discussed further in the section on grading practices. To reiterate, the consensus view at this time is that the association between SET ratings and objective measures of learning is essentially zero.

4. **Student ratings provide useful information.**

While the AIC is aware of the controversy over the use of SETs in evaluating teaching effectiveness, we believe that their potential benefits support their continued use. When properly used, they provide one source of data that cannot be obtained in any other way and can be valuable as a tool that helps faculty reflect on and improve their teaching. Moreover, the existence of the SFF signals to the students that Temple cares about teaching and that it takes their opinions seriously. It is equally vital for the University administration to let the faculty know that it

values teaching and that it considers assessment of teacher quality and continuous improvement to be critical in all evaluations.

5. Student ratings are not the only way to evaluate faculty teaching.

While SETs are one source of data, it is never appropriate to evaluate an instructor's teaching on these data alone. This has long been recognized but is now mandated for those instructors in colleges under the TAUP contract (please consult the current TU-TAUP contract for details on the use of SFFs). The Center for the Advancement of Teaching (CAT) has expertise on this topic and will both assist faculty with effectively using SFF data to improve teaching, and help deans and chairs to identify appropriate additional methods of teaching evaluation if asked.

Advice for instructors on how to make the best use of SFF data



All instructors have access to their SFF data at the end of each course. Importantly, SFFs are administered because of how these data are used and what they indicate. The following are some recommendations about how to make the best use of these data. Each recommendation is presented in the form of a question.

1. Do I need to only rely on the University's SFF for feedback?

All of the recommendations presented below focus on the use of the SFF report that an instructor receives at the end of the course. However, some of the most useful feedback can be obtained during the course, especially if this feedback is obtained no later than the midpoint when there is still time to make adjustments to your course. Soliciting this feedback can be quite simple and does not need to take much time. Basically, it is valuable to ask your students to anonymously tell you how they think the course is going. The Center for the Advancement of Teaching (CAT) recommends that you ask students these simple questions: What should I continue doing in this course because it helps you to learn? What should I stop doing because it doesn't help you to learn? What should I start doing because you believe it will help you to learn? What can you do differently to help yourself to learn? These responses are for your use only and are not intended to be reviewed by anybody else. If possible, make clear to your students that you have read and heard their feedback (both positive and negative). You want your students to know that you valued the time they put into providing this feedback and that you will incorporate suggestions that you believe are appropriate to your course and will enhance student learning.

2. Why should I add my own questions to the SFF?

The core questions on the SFF are general as they are intended to be used in all courses in the University. While these questions provide useful feedback, they do not give you specific information about the way you designed and taught your course. The questions you can add to the SFF allow you to obtain your students' perceptions about aspects of your course that are unique to your teaching or to your course. In the current system, the responses to these questions can be viewed only by you unless you decide to provide them to others. Some examples of items you may add include:

- If you recently changed your course text, consider adding: *The instructional materials for this course (books, handouts, etc.) were valuable in helping me learn.*
- If you used a new technology resource in your course, consider adding: *The use of educational technology helped my learning.*
- If you changed the way you provided feedback on writing assignments, consider adding: *The feedback I received in this class helped improve my learning.*

3. How do I know if my ratings are "good" or "bad"?

SFF reports include some comparative data. Included with the ratings is a comparison to typical University and School ratings for the same question on your course. However, instructors should always look at the SFF form in total, not piecemeal. In general, a large majority of all instructors obtain means of 4.0 or higher on the questions. If your mean is considerably below this (say 3.0 or lower) and if a majority of students indicate that they "Strongly Disagree" or "Disagree" with a statement, then, in your students' perceptions, they do not believe you were doing

as they hoped in that area. As an instructor, you should ask yourself why you think they have this perception. What are you doing, or not doing, that has caused them to give you this rating?

4. Use ratings carefully from courses where only a few students complete the SFFs.

The current Temple system allows students to rate courses if the enrollment is five or more. However, the number of students who complete the SFFs might be smaller than this. In small-response courses, even one or two low scores can shift the mean lower, even though those students' views are not representative of the majority of students. In general, if fewer than 10 students complete the SFFs for the course, the mean rating is not very useful. This is not a hard cutoff – the smaller the absolute number, the less weight should be placed on the reported average. Data obtained from classes where there is a low response rate are also not very informative (see #8, below).

5. What should I do if my ratings are consistently low?

This is a follow-up to the point made above. Any instructor can obtain low ratings on occasion. If, however, your ratings are consistently low (3.5 or less on most questions over several semesters and in different courses), then your students do not perceive you as a good instructor and you should do something about it. It should be emphasized that high SFF ratings are not the goal of instruction: the goal is student learning. Still, students who perceive that their instructor is not meeting their expectations may lose interest and engagement in the course. A suggested course of action would be this: start with a peer that you feel comfortable with and ask that person to attend your course. That person might give you feedback on why your students are rating you poorly. Another course of action is to consult with the Center for the Advancement of Teaching (CAT). One of their specific goals is to assist instructors in their teaching role. The Center is staffed by professionals with the necessary expertise. This help will be private and individualized and is a resource you should use.

6. Should I worry that if I give low grades, I will get low SFFs?

This is one of the major controversies about Student Evaluations of Teaching and has been the subject of a substantial body of research. While it is true that there is a positive correlation between grades and student evaluations, the correlation is not as high as many instructors believe. The research seems to indicate that what most students evaluate is the fairness of the grade they received and the clarity on which the evaluation was based. If your assignments and assessments are clearly linked to the course goals, and if you are clear about the way you give grades, then your students will be less likely to give you poor ratings if they receive a low grade.

7. How should I address negative comments that I think are unfair?

Almost all instructors receive negative comments from students on occasion. While negative comments should not be ignored, what you should look for are patterns and not pay too much attention to a single negative comment no matter how hurtful. However, if several students make the same or similar negative comments, then they are telling you that in their opinion there is something you are doing that they perceive as not facilitating their learning. This perspective should be addressed by at least acknowledging its existence and then attempting to understand on what basis the comment is made and whether there might be remedies for the perceived problem.

8. What should I do if a low percentage of my students complete the SFFs?

At present, the University average percentage of students who complete the SFFs is around 60% during most fall and spring semesters. If your average is consistently lower than this, there are a few things you can do. Some instructors provide time, typically at the end of the last class, for the students to complete the SFFs using their phones or other devices. This can be successful although it doesn't always work and may not fit the way your course is taught. The literature is very clear on one thing in this area: instructors who tell their students that the SFFs are important and that they will take the responses seriously obtain higher percentage return rates. It is critical that you tell your students that completing the SFFs is important to you and, if possible, give examples of how you used feedback to enhance your teaching and/or the course.

Guidelines for the use of SFF data to evaluate instructors



Faculty in evaluative roles, including deans and chairs, are rarely provided guidance on interpreting other faculty members' SFFs. In the absence of research-based guidance, these faculty may end up relying on their own experiences, biases, and possibly erroneous information. The guidelines presented below are intended to make this evaluative process fairer and more informed by research.

1. The most important thing is to use the SFF forms holistically as much as possible.

Myriad factors affect SFF scores on a particular form, and there is even more variability and less reliability in a single question on the form. As much as possible, evaluators should look for patterns that are replicated across multiple SFFs. Evaluators should also attempt to figure out reasonable explanations, based in research, for why those patterns might exist. Deviations up and down are to be expected as part of the normal variation across classes, semesters, and years, particularly if the number of respondents on a given SFF form is low.

2. An instructor's complete set of student ratings should be considered.

In general, ratings across the various items are similar, but the nuance gained by looking at each item can be valuable in certain circumstances. It may seem self-evident that the questions ask different things and that students will rate each question independently of the others. That is, there is no logical reason for a high rating on Q1 to imply a high rating on Q3. However, perhaps counterintuitively, extensive research and Temple experience finds that students tend to rate instructors very similarly across questions (no matter what the questions are). Thus, exercise caution in interpreting any particular question or subset of questions in isolation. The exception is if a particular question seems to break the pattern of all the other ones by a significant amount (e.g., everything else is a 5, but one question is a 3, or everything is a 3, but one question averages a 4.5). That type of pattern bears further investigation. As a general rule, the totality of the questions should be used in evaluation.

3. It is not appropriate to use a subset of questions to assess teaching adequacy.

This is a follow-up to Guidelines 1 & 2. Avoid using scores from two or three SFF questions that are viewed as "more important" or creating a combined score from only those questions (occasionally in combination with an established cut-point) for decisions such as merit or contract renewal. This is **not** an appropriate use of SFF data. First, small differences in scores are not meaningful. Second, the SFF forms do not measure inherent teaching or teaching ability; rather, they aggregate student perception data unique to the instructor/class combination for the semester. Third, taking subsets of questions decreases the reliability of measurement. As mentioned above, responses should be viewed holistically.

4. It is not appropriate to use the data from a single semester or a single course.

When evaluating an instructor's teaching, it is always better to use data across multiple semesters and courses. While this is not always possible, multiple data sources are always better than a single assessment.

5. Small differences in ratings are common and not necessarily meaningful.

At Temple, the average score on most items on the SFF is in the 4.0 to 4.3 (out of 5) range. For a variety of statistical reasons, small differences between scores are not likely to be meaningful, particularly at the top end of the distribution. These scores are not normally distributed and are highly skewed. As such, care must be taken in over-interpreting small differences. The difference between a mean of 4.3 and 4.5, for example, is not meaningful.

6. Be cautious in using anomalous ratings.

An anomalous rating for an entire course likely had some identifiable cause behind it but is unlikely to be a good representation of what students in general would think of that course/professor combination. Small anomalous ratings within a given SFF form are rarely meaningful, but very large deviations on a single question of a form should spark an attempt to determine whether the result was a random anomaly or had a reason behind it. One way or another, it is always better to look for patterns in an instructor's rating over time or across different course types. Every instructor receives an occasional low rating. While one unsatisfactory set of ratings should not be ignored, they should also not be over-interpreted. It is particularly important to keep in mind that an anomalous negative rating might be due to an instructor having been assigned, or having volunteered, for a particularly difficult or undesirable teaching assignment, a new teaching assignment, or a late assignment. Over-interpreting one unsatisfactory set of ratings may also discourage innovation in teaching as faculty might be rightly concerned about detrimental effects on their evaluations.

7. Use ratings carefully from courses where only a few students complete the SFFs.

The current Temple system allows students to rate courses if the enrollment is five or more. However, the number of students who complete the SFFs might be smaller than this. In small-response courses, even one or two low scores can shift the mean lower, even though those students' views are not representative of the majority of students. In general, if fewer than 10 students complete the SFFs for the course, the mean rating is not very useful because one or two students' responses can have a significant effect. This is not a hard cutoff – the smaller the absolute number, the less weight should be placed on the reported average. Data obtained from classes where there is a low response rate are not very informative and will exhibit greater variability.

8. Discuss low response rates with instructors as they might indicate a lack of commitment by the students.

This recommendation follows from the one above. Temple has an interest in having students provide feedback. One major reason for a low response rate is because the instructor did not provide time in class to complete the SFF. Of course, another possibility is that students were simply uninterested in providing feedback. Lack of interest in providing feedback might be an indication that students were not deeply engaged in the course but also might mean that they were mostly satisfied with the course and the instructor. There are other factors that could affect response rate; for example, research shows that online courses receive lower response rates than in-person courses, and students are unlikely to be as engaged in a first-year required course than an upper-level elective. When discussing SFF data with instructors, a low response rate is something to mention. Encourage instructors to share with their students that they value their feedback. Giving time in class signals to students that the instructor values their feedback and is willing to give time for them to complete the SFF. Instructors should also share how they have used student feedback in the past to make adjustments to the course or their teaching. Other strategies to increase response rates can be found [here](#).

9. Avoid comparing faculty to each other.

Student rating instruments are not designed to gather comparative data about instructors. The purpose of these instruments is to gain an overall sense of students' perceptions of a single instructor teaching a particular course (or part of a course) to a specific group of students. As mentioned above, SFF forms do not directly measure the main outcomes upon which instructors are compared (for example, SFF forms measure neither teaching ability nor student learning). Comparisons should be sparing and limited to what can be validly defended. For example, in a multi-section course, one might use SFF scores as a general indicator of student satisfaction across sections. However, extreme care should be taken to ensure that the comparison being made actually applies. For example, even when teaching sections of the same course, sections may not be comparable (e.g., the MWF 8 am section may or may not be comparable to the TR 2 pm section) and instructors who do not fit the common stereotype in a specific field may not be comparable (e.g., female instructors in male dominated fields).

10. Always read the student comments.

This is another recommendation that seems so obvious that it does not need to be stated, but it is important enough to include. The Temple SFF contains several open-ended questions that students are asked to complete. These student comments offer valuable information that cannot be provided by numerical ratings alone. There is a commonly held belief that only students with more extreme views, both positive and negative, respond to these open-ended comments. While the literature does not strongly support this belief, it is the case that the students who provide comments are the ones who are committed enough to take the time to do so. These comments often provide the most useful information for understanding the ratings.

11. Focus on the most common comments rather than emphasizing one or a few atypical ones.

This recommendation follows the one above and offers some cautions about the use of student comments to evaluate teaching. When evaluating an instructor's teaching by reading the comments, common themes should be emphasized. It is sometimes the case in a set of comments to find a few that differ from the majority. Strongly negative comments should not be ignored, but they also should not be given more weight than the views of most students. This is particularly crucial when evaluating the ratings of non-majority faculty where this problem is more common. It is also important to understand whether comments form a pattern across courses and over time or are just a result of a single course or class dynamic.

12. Contradictory written comments are not unusual.

This is an extension of the previous recommendation, but it is less focused on negative comments. As mentioned above, the best use of student comments is to search for themes. It is not uncommon, however, to find completely contradictory perceptions in these comments: some students think the textbook is great, others hate it; some students want more group work, others want less. The fact that these contradictions exist is not necessarily a sign of poor teaching. Remember that student feedback data are perceptions, and perceptions may vary.

13. Use an instructor's grading practices as one context for reviewing SFF data.

This recommendation focuses on one of the most controversial issues in student evaluation of teaching: the relationship between grades and evaluations. Perhaps the most common criticism of these evaluations is that faculty can "buy" good evaluations by giving high grades. The literature is very clear that grades and evaluations are positively correlated, and that ratings are affected by a student's expected grade in a course. While the correlation is lower than many believe, it is still one of the strongest effects in the research literature. The presence of this effect is problematic. With this in mind, one suggestion is to examine an instructor's grades when examining the instructor's SFFs. Keeping in mind that grades in a particular course can be higher or lower than normal for very good reasons (e.g., a particular group of students is unusually unprepared for the work in the course), a pattern of very high grades, across semesters and in different courses, is something that is worth discussing.

14. Always use multiple measures to assess instruction.

This is good practice and part of holistic assessment. SFF data simply do not provide information about many elements that are highly relevant to whether someone is a good instructor. In addition, where applicable, the TU-TAUP contract requires that SFF data cannot be the sole way to assess instruction. As mentioned, the Center for the Advancement of Teaching will provide assistance to any college or department to help develop additional assessment processes.

Advice to chairs and deans on how to speak to faculty about their SFFs



Any form of teaching evaluation is most useful if it is used as a basis for discussion with faculty members. Much time and energy are put into developing and administering evaluations. Yet, sometimes not enough time is put into examining, understanding and discussing the evaluations with faculty to aid them in interpreting the results and/or discussing ways to improve teaching. The guidelines presented above are intended for a wide variety of people in evaluative roles. In practice, most of the feedback that is provided to instructors comes through their department chairs or deans. As such, the AIC felt that some suggestions intended for chairs and deans would be valuable. These suggestions are presented below.

When meeting with the faculty in your department to discuss student feedback, some best practices can facilitate a productive discussion among colleagues and an opportunity to reflect on how to improve student perceptions of teaching.

Planning for Teaching Discussions

- 1. Set aside uninterrupted time to discuss SFF results with all faculty members.**
It is important to talk to all faculty, not just those that you might consider problematic. It is just as important to spend time discussing what went right as it is to discuss what went awry.
- 2. Ask each faculty member to read their own SFFs before meeting with you.**
Always ask the instructor to come prepared to discuss the larger patterns of positive and negative comments that they see in the feedback.
- 3. Read the SFFs carefully.**
When evaluating an instructor's teaching, always look for patterns of feedback and choose at most three areas of improvement and three areas of strength that you wish to focus on in the meeting.

Suggestions for Conducting the Meeting

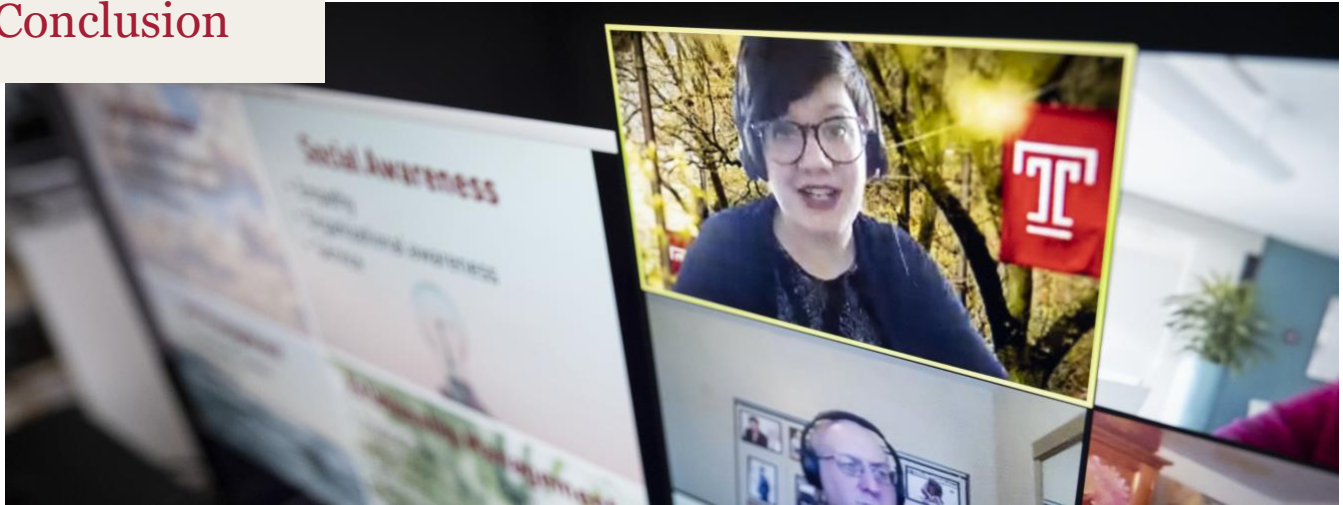
- 4. Ask questions first.**
When you meet with a faculty member, start the discussion by asking the faculty member to share what resonated with them in the feedback. You might ask them what they think of the feedback, why students might have responded in that way to the teaching practice, and whether they are considering any changes based on the feedback. A series of questions instead of statements will lead to more reflection on the part of the faculty member and open the way for a productive discussion.

If necessary...

5. **Discuss missed topics after the faculty member offers their viewpoint.**
After the faculty member has gone through their self-evaluation, bring up any areas of improvement or strength that you marked as areas of focus that have not been discussed.
6. **Offer your own constructive ideas in the form of questions.**
“Do you think it would work if...?” If you are well known as a good teacher, then using your own experience can be powerful: “I have often found that if I do X, students respond well. Do you think that would work for you?” If not, phrasing more generically is best (e.g., “It has been found that when instructors do X, students respond well. Do you think that might work for you?”). Make sure to make it clear that this is a process of exploration and brainstorming ideas for improvement instead of a critique.
7. **Develop an action plan.**
Ask the faculty member to decide on two or three concrete steps they will take to improve their teaching or the course.
8. **Dealing with recalcitrant faculty.**
If the faculty member is resisting the idea of change and improvement, point out that it is important for every faculty member to contribute to an environment of positive engagement in order for the department (school) to continue to thrive. Remind them that there are resources (such as the Center for the Advancement of Teaching and Temple’s Institutional Diversity, Equity, Advocacy and Leadership—IDEAL) that can help them think through challenges. Note that if SFF feedback is indicative of faculty behavior that demonstrates a lack of sensitivity to the diversity of students in the class (e.g., race, ethnicity, national origin, gender, sexual identity, disability, or political viewpoint), make a more concrete plan in writing for the faculty member’s improvement, refer them to your dean’s office, and set up consultations with the CAT and/or IDEAL. You must insist that this behavior is not acceptable and must be remediated.

The suggestions presented in this document are intended to provide some guidance to deans, department chairs or anyone in a position that requires providing feedback to instructors about their teaching. If the student feedback is generally positive, this conversation is not problematic. If, however, the feedback is largely negative, and if similar feedback has occurred across several semesters, the task is not easy. How do you tell a long-serving faculty member who has had poor student ratings for years that those ratings are no longer acceptable? Angela Linse suggests opening the conversation with a statement like this: “It may have been sufficient in the past to receive these kinds of ratings, but things have changed and students expect more now. The University has invested resources to help you take the next steps to improve your teaching. For example, ...”. At this point the Center for the Advancement of Teaching can be mentioned, or, if the college or department has created their own professional development activities, these can be mentioned. It is important to try and keep the conversation as positive as possible but to emphasize that improvement is necessary and that it will be monitored.

Conclusion



This document provides an overview of the best practices in the use and interpretation of these ratings by instructors, promotion and tenure committees, and others who evaluate an instructor's teaching. It is intended as a resource to make the evaluative process more consistent across departments and schools/colleges, promote data-informed conversations about use of teaching evaluations, and offer strategies for having productive conversations about SFF data with faculty.

The conversation and research on student evaluation of teaching (SETs) is dynamic and changing, with new research continuing to emerge. This set of recommendations will be reviewed and updated to reflect new research as well as the experiences of those using these recommendations. We hope that this document is useful. If you have recommendations for further enhancement of this document, please share your ideas with sff@temple.edu.

References¹

- Abrami, P. C. (2001). Improving judgments about teaching effectiveness using teacher rating forms. *New Directions for Institutional Research*, 109, 59–87.
- Abrami, P. C., d'Appolonia, S., & Cohen, P. A. (1990). The validity of student ratings of instruction: What we know and what we don't. *Journal of Educational Psychology*, 82(2), 219–231.
- Abrami, P. C., Dickens, W. J., Perry, R. P., & Leventhal, L. (1980). Do teacher standards for assigning grades affect student evaluations of instruction? *Journal of Educational Psychology*, 72, 107–118.
- Aleamoni, L. M. (1999). Student rating myths versus research facts: An update. *Journal of Personnel Evaluation in Education*, 13(2), 153–166.
- Ardalan, A., Ardalan, R., Coppage, S., & Crouch, W. (2007). A comparison of student feedback obtained through paper-based and web-based surveys of faculty teaching. *British Journal of Educational Technology*, 38(6), 1085–1101.
- Anderson, K. J., & Smith, G. (2005). Students' preconceptions of professors: Benefits and barriers according to ethnicity and gender. *Hispanic Journal of Behavioral Sciences*, 27(2), 184–201.
- Arreola, R. A. (2007). *Developing a comprehensive faculty evaluation system* (3rd ed.). Anker.
- Bachen, C. M., McLoughlin, M. M., & Garcia, S. S. (1999). Assessing the role of gender in college students' evaluations of faculty. *Communication Education*, 48(3), 193–210.
- Barkley, E. F. (2010). *Student engagement techniques: A handbook for college faculty*. Jossey-Bass.
- Barre, B. (2015). Academic blogging and student evaluation click bait: A follow-up. *Reflections on Teaching and Learning, the CTE Blog*. <http://cte.rice.edu/blogarchive/2015/07/28/student-evaluationsfollowup>.
- Basow, S. A. (1995). Student evaluations of college professors: When gender matters. *Journal of Educational Psychology*, 87, 656–665.
- Benton, S. L. & Cashin, W. E. (2011). Student ratings of teaching: A summary of research and literature - IDEA Paper No. 50. *IDEA Center*. https://www.ideaedu.org/idea_papers/student-ratings-of-teaching-a-summary-of-research-and-literature/
- Benton, S. L., Guo, M., Li, D., & Gross, A. (2013, April). *Student ratings, teacher standards, and critical thinking skills* [Paper presentation]. American Educational Research Association Annual Meeting, San Francisco, CA.
- Benton, S. L. & Li, D. (2015). Response to "a better way to evaluate undergraduate teaching" - IDEA Editorial Note #1. *IDEA Center*. https://www.ideaedu.org/Portals/0/Uploads/Documents/A_Better_Way_to_Evaluate.pdf
- Benton, S. L., Webster, R., Gross, A. B., & Pallett, W. H. (2010). An analysis of IDEA student ratings of instruction using paper versus online survey methods, 2002–2008 data - IDEA technical report no. 16. *Idea Center*. <http://ideaedu.org/wp-content/uploads/2014/11/techreport-16.pdf>
- Berk, R. A. (2013). *Top 10 flashpoints in student ratings and the evaluation of teaching: What faculty administrators must know to protect themselves in employment decisions*. Stylus.
- Berk, R. A. (2012). Top 20 strategies to increase the online response rates of student rating scales. *International Journal of Technology in Teaching and Learning*, 8(2), 98–107.
- Berk, R. A. (2006). *Thirteen strategies to measure college teaching: A consumer's guide to rating scale construction, assessment, and decision making for faculty, administrators, and clinicians*. Stylus.
- Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17(1), 48–62.
- Boice, R. (2001). *Advice for new faculty members: Nihil nimis*. Allyn and Bacon.
- Boyer, E. L. (1990). *Scholarship reconsidered: Priorities of the professoriate*. Carnegie Foundation for the Advancement of Teaching.
- Bragaa, M., Paccagnellab, M., & Pellizzaric, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, 41, 71–88.
- Braskamp, L. A., Brandenburg, D. C., & Ory, J. C. (1984). *Evaluating teaching effectiveness: A practical guide*. Sage.
- Braskamp, L. A., Ory, J. C., & Pieper, D. M. (1981). Student written comments: Dimensions of instructional quality. *Journal of Educational Psychology*, 73(1), 65–70.
- Brinko, K. T. (1991). The interactions of teaching improvement. *New Directions for Teaching and Learning*, 48, 21–37.
- Cashin, W. E. (2003). Evaluating college and university teaching: reflections of a practitioner. In Smart, J. C. (Ed.), *Higher education: Handbook of theory and research* (pp. 531–593). Kluwer Academic.
- Cashin, W. E. (1999). Student ratings of teaching: uses and misuses. In Seldin, P. (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 25–44). Anker.
- Cashin, W. E. (1996). Developing an effective faculty evaluation system - IDEA paper no. 33. *IDEA Center*. http://ideaedu.org/wp-content/uploads/2014/11/Idea_Paper_33.pdf
- Cashin, W. E. (1995). Student ratings of teaching: The research revisited - IDEA paper no. 32. *IDEA Center*. <http://files.eric.ed.gov/fulltext/ED402338.pdf>
- Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *Journal of Higher Education*, 71(1), 17–33.
- Chism, N. V. (2007). *Peer review of teaching: A sourcebook*. Anker.
- Cox, M. D. (2004). Introduction to faculty learning communities. *New Directions for Teaching and Learning*, 97, 5–23.
- d'Appolonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist*, 52(11), 1198–1208.
- Davis, D. J. (2010). The experiences of marginalized academics and understanding the majority: Implications for institutional policy and practice. *International Journal of Learning*, 17(6), 355–364
- Dommeier, C. J., Baum, P., Hanna, R. W. & Chapman, K. S. (2004). Gathering faculty teaching evaluations by in-class and online surveys: Their effects on response rates and evaluations. *Assessment & Evaluation in Higher Education*, 29(5), 611–623.
- Eiszler, C. F. (2002). College students' evaluations of teaching and grade inflation. *Research in Higher Education*, 43(4), 483–501.
- Fairweather, J. S. (2002). The ultimate faculty evaluation: Promotion and tenure decisions. *New Directions for Institutional Research*, 114, 97–108.
- Feldman, K. A. (2007). Identifying exemplary teachers and teaching: evidence from student ratings. In Perry, R. P., & Smart, J. C. (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 93–95). Springer.
- Feldman, K. A. (1993). College students' views of male and female college teachers: Part II—Evidence from students' evaluations of their classroom teachers. *Research in Higher Education*, 34(2), 151–211.
- Feldman, K. A. (1992). College students' views of male and female college teachers: Part I— Evidence from the social laboratory and experiments. *Research in Higher Education*, 33(3), 317–375.
- Feldman, K. A. (1989). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students, colleagues, administrators, and external (neutral) observers. *Research in Higher Education*, 30(2), 137–189.
- Feldman, K. A. (1976). Grades and college students' evaluations of their courses and teachers. *Research in Higher Education*, 4, 69–111.
- Franklin, J. (2001). Interpreting the numbers: Using a narrative to help others read student evaluations of your teaching accurately. *New Directions for Teaching and Learning*, 87, 85–100.
- Franklin, J., & Berman, E. (1998). Using student written comments in evaluating teaching. *Instructional Evaluation and Faculty Development*, 18(1) [text version in possession of the author has no page numbers].
- Franklin, J., & Theall, M. (1994, April 7). *Student ratings of instruction and sex differences revisited* [Paper presentation]. American Educational Research Association Annual Meeting, New Orleans, LA.
- Franklin, J. L., & Theall, M. (1991, April 7). *Grade inflation and student ratings: a closer look* [Paper

¹ Most references are from Angela Linse's Report to the Penn State Faculty Senate

- presentation]. American Educational Research Association Annual Meeting, Chicago, IL.
- Galguera, T. (1998). Students' attitudes towards teachers' ethnicity, bilinguality, and gender. *Hispanic Journal of Behavioral Sciences, 20*(4), 411–429.
- Geis, G. L. (1991). The moment of truth: Feeding back information about teaching. *New Directions for Teaching and Learning, 48*, 7–19.
- Gigliotti, R. J., & Buchtel, F. S. (1990). Attributional bias and course evaluations. *Journal of Educational Psychology, 82*(2), 341–351.
- Gilroy, M. (2007). Bias in student evaluations of faculty? *The Hispanic Outlook in Higher Education 17*(19), 26–27.
- Glassick, C. E., Huber, M. T., & Maeroff, G. I. (1997). *Scholarship assessed: Evaluation of the Professorate*. Jossey-Bass.
- Greenwald, A. G., & Gillmore, G. M. (1997). Grading lenience is a removable contaminant of student ratings. *American Psychologist, 52*(11), 1209–1217.
- Gunsalus, C.K. (2006). *The college administrator's survival guide*. Harvard University Press.
- Gutgold, N. D., & Linse, A. R. (2016). *Women in the academy: Learning from our diverse career pathways*. Lexington.
- Hancock, G. R., Shannon, D. M., & Trentham, L. L. (1993). Student and teacher gender in ratings of university faculty: results from five colleges of study. *Journal of Personnel Evaluation in Education, 6*(3), 235–248.
- Hardy, N. (2003). Online ratings: fact and fiction. *New Directions for Teaching and Learning, 96*, 31–38.
- Hativa, N. (2013a). *Student ratings of instruction: A practical approach to designing, operating, and reporting*. Oron Publications.
- Hativa, N. (2013b). *Student ratings of instruction: Recognizing effective teaching*. Oron Publications.
- Hendrix, K. J. (1998). Student perception of the influence of race on professor credibility. *Journal of Black Studies, 28*(6), 738–763.
- Huber, M. T. (2002). Faculty evaluation and the development of academic careers. *New Directions for Institutional Research, 114*, 73–83.
- Husbands, C. T. (1997). Variations in students' evaluations of teachers' lecturing in different courses on which they lecture: A study at the London School of Economics and Political Science. *Higher Education, 33*, 51–70.
- IDEA Research Note 1. (2003). The "excellent teacher" item. *The IDEA Center*.
- Johnson, T. D. (2003). Online student ratings: will students respond? *New Directions for Teaching and Learning, 96*, 49–59.
- Kaplan, M. (2014). Release of course evaluations to students, policies of University of Michigan peer institutions. *Center for Research on Learning and Teaching, University of Michigan*.
- Kulik, J. A. (2001). Student ratings: Validity, utility, and controversy. *New Directions for Institutional Research, 109*, 9–25.
- Lazos, S. R. (2011) Are student teaching evaluations holding back women and minorities? The perils of "doing" gender and race in the classroom. In, Gutiérrez y Muhs, G., Niemann, Y. F., González, C. G., & Harris, A. P. (Eds.), *Presumed incompetent: The intersections of race and class for women in academia* (pp. 164–185). University Press of Colorado.
- Lewis, K. G. (2001). Making sense of student written comments. *New Directions for Teaching and Learning, 87*, 25–32.
- Lewis, K. G. (1991). Gathering data for the improvement of teaching: What do I need and how do I get it? *New Directions for Teaching and Learning, 48*, 65–82.
- Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation 54*, 94–106. <http://dx.doi.org/10.1016/j.stueduc.2016.12.004>
- Linse, A. R. (2010). *Analysis of online SRTE data from select semesters (2009–2010)*. Schreyer Institute for Teaching Excellence, The Pennsylvania State University.
- Linse, A. R., & Xie, H. (2011). *Student ratings of teaching effectiveness: Analysis of data from common courses from select semesters (2009–2010)*. Schreyer Institute for Teaching Excellence, The Pennsylvania State University.
- MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education, 40*, 291–303.
- Marincovich, M. (1999). Using student feedback to improve teaching. In Seldin, P. (Ed.), *Changing practices in evaluating teaching: a practical guide to improved faculty performance and promotion/tenure decisions* (pp. 45–69). Anker.
- Marsh, H. W. (2007). Students' evaluations of university teaching: A multidimensional perspective. In Perry, R. P., & Smart, J. C. *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp.319–384). Springer.
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research 11*(3), 253–388.
- 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. (1982a). Factors affecting students' evaluations of the same course taught by the same instructor on different occasions. *American Educational Research Journal, 19*(4), 485–497.
- Marsh, H. W. (1982b). Validity of students' evaluations of college teaching: A multitrait-multimethod analysis. *Journal of Educational Psychology, 74*, 264–279.
- Marsh, H. W. (1980). Research on students' evaluations of teaching effectiveness. *Instructional Evaluation, 4*(5), 5–13.
- Marsh, H. W. & Dunkin, M. J. (1992). Students' evaluations of university teaching: A multidimensional perspective. In Smart, J. C. (Ed.), *Higher education: Handbook of theory and research* (Vol. 8, pp. 143–233). Agathon.
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist, 52*, 1187–1197.
- McGhee, D. E., & Lowell, N. (2003). Psychometric properties of student ratings of instruction in online and on-campus courses. *New Directions for Teaching and Learning, 96*, 39–48.
- McKeachie, W. J. (1997). Student ratings: the validity of use. *American Psychologist, 52*(11), 1218–1225.
- McKeachie, W. J. (1990). Research on college teaching: the historical background, *Journal of Educational Psychology, 82*(2), 189–200.
- McKeachie, W. J. (1979). Student ratings of faculty: A reprise. *Academe, 65*(6), 384–397.
- Miller, J. E. & Seldin, P. (2014) Changing practices in faculty evaluation: Can better evaluation make a difference? *Academe, 100*(3), 35–38.
- National Academies. (2006). *Beyond bias and barriers: Fulfilling the potential of women in academic science and engineering*. Committee on Maximizing the Potential of Women in Academic Science and Engineering, and Committee on Science, Engineering and Public Policy.
- Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment and Evaluation in Higher Education, 33*(3), 301–314.
- Ory, J. C. (2001). Faculty thoughts and concerns about student ratings. *New Directions for Teaching and Learning, 87*, 3–15.
- Ory, J. C., Braskamp, L. A., & Pieper, D. M. (1980). Congruency of student evaluative information collected by three methods. *Journal of Educational Psychology, 72*, 181–185.
- Ory, J. C., & Ryan, K. (2001). How do student ratings measure up to a new validity framework? *New Directions for Institutional Research, 109*, 27–44.
- Ouellett, M. L. (2010). Overview of faculty development: History and choices. In Gillespie, K., & Robertson, D. (Eds.), *A guide to faculty development* (2nd ed., pp. 3–20). Jossey-Bass.
- Pallett, W. H. (2006). Uses and abuses of student ratings. In Seldin, P. (Ed.), *Evaluating faculty performance*, (pp. 50–65). Anker.
- Remmers, H. H. (1933). Learning, effort, and attitudes as affected by three methods of instruction in elementary psychology. *Purdue University Studies in Higher Education* (Monograph No. 21).
- Remmers, H. H., & Brandenburg, G. C. (1927). Experimental Data on the Purdue Rating Scale for Instruction. *Educational Administration and Supervision, 13*, 519–527.
- Reid, L. D. (2010) The role of perceived race and gender in the evaluation of college teaching on RateMyProfessors.com. *Journal of Diversity in Higher Education, 3*(3), 137–152.

- Ryalls, K., Benton, S., Barr, J., & Li, D. (2016). Response to "bias against female instructors" - IDEA Editorial Notes. *IDEA Center*.
- Seldin, P. (1999). *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions*. Anker.
- Sinclair, L., & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me, *Personality and Social Psychology Bulletin*, 26(11), 1329–1342.
- Smith, B. P., (2009). Student ratings of teaching effectiveness for faculty groups based on race and gender. *Education*, 129(4), 615–624.
- Smith B. P. (2007) Student ratings of teaching effectiveness: An analysis of end-of-course faculty evaluations. *College Student Journal*, 41(4), 788–800.
- Smith, B. P., & Hawkins, B. (2011) Examining student evaluations of black college faculty: Does race matter? *The Journal of Negro Education*, 80(2), 149–162.
- Smith, B. P., & Johnson-Bailey, J. (2011/2012). Implications for non-white women in the academy. *The Negro Educational Review*, 62 & 63 (1–4), 115–140.
- Soderberg, L. O. (1985). Dominance of research and publication: An unrelenting tyranny. *College Teaching*, 33, 169–172.
- Sorcinelli, M. D., & Austin, A. E. (2006). Developing faculty for new roles and changing expectations. *Effective Practices for Academic Leaders*, 1(11), 1–16.
- Sorcinelli, M. D., Austin, A. E., Eddy, P. L., & Beach, A. L. (2006). *Creating the future of faculty development: Learning from the past, understanding the present*. Anker.
- Sorenson, D. L, & Reiner, C. (2003). Charting the uncharted seas of online student ratings of instruction. *New Directions for Teaching and Learning*, 96, 1–24.
- Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluations of teaching: The state of the arts. *Review of Educational Research*, 83(4): 598–642.
- Stowell, J. R., Addison, W. E., & Smith, J. L. (2012). Comparison of online and classroom-based student evaluations of instruction. *Assessment and Evaluation in Higher Education*, 37(4), 465–473.
- Street, S., Kimmel, E., & Kromrey, J.D. (1996). Gender role preferences and perceptions of university students, faculty, and administrators. *Research in Higher Education* 37(5), 615–632.
- Stumpf, S. A., & Freedman, R. D. (1979). Expected grade covariation with student ratings of instruction: Individual versus class effects. *Journal of Educational Psychology*, 71, 293–302.
- Svinicki, M. D. (2001). Encouraging your students to give feedback. *New Directions for Teaching and Learning*, 87, 17–24. Jossey-Bass.
- Theall, M., & Franklin, J. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction? *New Directions for Institutional Research*, 109, 45–56.
- Theall, M., & Franklin, J. (2000). Creating responsive student ratings systems to improve evaluation practice. *New Directions for Teaching and Learning*, 83, 95–107.
- Theall, M., & Franklin, J. (1990). Editors Notes. *New Directions for Teaching and Learning*, 43, 1–14.
- Venette, S., Sellnow, D., & McIntyre, K. (2010). Charting new territory: Assessing the online frontier of student ratings of instruction. *Assessment & Evaluation in Higher Education*, 35(1), 101–115.
- Webster, R. J., Benton, S., & Gross, A. (2010, April 3–May 4). *Online versus paper survey delivery of college student ratings of instruction* [Paper presentation]. American Educational Research Association Annual Meeting, Denver, CO.
- Wilson, R. C. (1986). Improving faculty teaching: Effective use of student evaluations and consultants. *The Journal of Higher Education*, 57(2), 196–211.
- Zakrajsek, T. D. (2010). Important skills and knowledge. In Gillespie, K., & Robertson, D. (Eds.), *A guide to faculty development* (2nd ed., pp. 83–98). Jossey-Bass.