Teaching in Higher Education

Student evaluations of teaching: perceptions of faculty based on gender, position, and rank

Lori R. Kogan [a] , Regina Schoenfeld-Tacher [a] & Peter W. Hellyer [a]

[a] Department of Clinical Sciences, Colorado State University,
Campus Delivery, Fort Collins, CO, USA

Available online: 20 Oct 2010

PLEASE SCROLL DOWN FOR ARTICLE

Student evaluations of teaching: perceptions of faculty based on gender, position, and rank

Lori R. Kogan*, Regina Schoenfeld-Tacher and Peter W. Hellyer

Department of Clinical Sciences, Colorado State University, Campus Delivery, Fort Collins CO, USA

The current study explores the feelings and thoughts that faculty have about their student evaluations of teaching (SET). To assess the perceptions of SETs, all teaching faculty in one college at a western Land Grant University were asked to complete an anonymous online survey. The survey included demographic questions (i.e. gender; rank such as assistant, associate, and full professor; and positions like non-tenure track, tenure track, and tenured) as well as questions related to faculty's feelings while reading their SETs. While minimal differences were found in responses based on rank or position, several differences were found based on faculty gender. Overall, female faculty appear to be more negatively impacted by student evaluations than male faculty. These gender differences support previous research that suggests males and females receive and react differently to personal evaluation. Resultant suggestions include modifying surveys from anonymous to confidential and offering professional development training for faculty.

Keywords: student teaching evaluations; gender; assessment; faculty

Introduction

The topic of student evaluations of teaching (SET) has received much attention in terms of reliability, validity, and potential uses. The emotional impact of SETs on faculty, however, has been largely ignored. The current study attempts to explore the feelings and thoughts that faculty have about their student evaluations and if these differ for male and female faculty.

Higher education has long placed importance on evaluating professors to ensure they are teaching in ways most conducive to learning. Increased accountability for faculty efforts has been recommended by the Spellings report (United States Department of Education 2006) and SETs have traditionally been the primary source of assessing teaching in higher education (Weinberg, Hashimoto, and Fleisher 2009). Given near the end of the semester, SETs are an almost universally accepted method of gathering information on teaching (Zabaleta 2007). It has been assumed that students are in the best position to know whether the teaching they receive is adequate and whether they are learning (Clayson and Haley 1990).

This idea appears deceptively simple. Surveys are developed using five- or seven-point Likert scales (for example, strongly agree, agree, neutral, disagree and strongly

---

disagree) that ask students to reflect on aspects of instruction, course topics, and the instructor. After students complete the surveys, summary statistics are calculated and assessment of teaching has been completed. In reality, the process is far more complicated, and a tremendous amount of controversy surrounding student evaluations continues.

Traditional SETs share some similar characteristics. They are usually comprised of a series of open and closed questions about course content and teaching effectiveness with at least one question pertaining to overall teaching effectiveness. They usually encourage written comments about the content and instruction. The surveys are typically anonymous and given near the end of the semester without the presence of the instructor – either in paper or electronic format. After the surveys are completed, the responses are usually summarized across instructors, departments, and colleges and viewed as evidence of teaching effectiveness that is then used for personnel and professional decisions (Sproule 2000).

Much of the debate on student evaluations is not about whether evaluations are needed, but whether the current instruments are reliable and valid, and whether they should be used in high-stake decisions. Some argue that it only logically follows that students tend to learn more from good teachers, so if the process of student evaluations is valid, there should be a strong positive correlation between student learning and evaluations. This has not consistently, however, been the case. Therefore, others argue that even this seemingly simplistic view of teaching and learning is complicated with numerous external factors that are beyond an instructor's control (e.g. student ability and motivation, physical environment). So even in the best of worlds, the relationship between SETs and student learning can be far from a simple correlation (Cohen 1981).

Despite the controversy, SETs are seen by many as a valuable tool to improve teaching and student learning outcomes. SETs are also often used for critical decisions, such as retention, tenure, and promotion of faculty. The practice of using SETs to make important personnel decisions, where high evaluations come with awards and promotions, and low evaluations result in some type of punitive action, is questioned by many faculty.

It may be argued that few areas within academia have been researched and documented to the degree that student evaluations have been (Clayson 2009). Since the 1970s, the use of student evaluations began expanding dramatically, so that currently the vast majority of schools now include some form of student evaluations (Comm and Manthaisel 1998). Given their popularity, it follows that many researchers support SETs and find them valid and useful (Cohen 1981; McKeachie 1986; Ramsden 2003; Wilson, Lizzo, and Ramsden 1997). Marsh (1987), for example, reported that SETs are multidimensional, reliable, stable, a function of the instructor rather than the course, relatively unaffected by a host of variables thought to be potential biases, and overall, a useful tool for faculty, students, and administration. Aleamoni (1999) argued that SETs can be both reliable and valid, and useful as tools to enhance teaching, course content, and document instructional effectiveness. Baird (1987) has reported that evaluations reflect what is actually learned and Marsh (1977) has defended the validity of SETs based on the fact that past and present students agree on overall evaluations of courses and instructors. Yet, even those who attest to the positive aspects of SETs suggest that

they should be used in conjunction with other forms of teaching performance evaluation (Marsh 1984).

Others have questioned SETs usefulness (Wright 2006). For example, Olshavsky and Spreng (1995) found a weak correlation between student ratings of instructors and final examination scores and Dowell and Neal (1982) found no relationship between learning and student evaluations. Other researchers (Rodin and Rodin 1972; Yunker and Yunker 2003) found a negative correlation between learning and student evaluations. Dowell and Neal (1983) explained these findings by stating that, far from evaluating teaching effectiveness, SETs at best can be viewed as indices of consumer satisfaction.

The personality of the instructor, for example, has been found to impact teaching evaluations, regardless of actual amount of information learned (Sherman and Backburn 1975). In fact, even type of dress has been found to impact students' views of instructors. Sebastian and Bristow (2008) found that students rated female instructors in formal dress as less likeable when compared to female instructors in casual dress. Style of dress in this study did not, however, impact likeability for male instructors. Physical attractiveness can also impact SETs, as illustrated by Hamermesh and Parker (2003). Consistent with this theme, another study found that students could accurately predict student evaluations from watching video clips of a lecture with the sound turned off (Ambady and Rosenthal 1993). This speaks partially to the entertainment factor, in which entertaining presenters, even actors with no topical knowledge base, are rated high on teaching evaluations (Costin, Greenough, and Menges 1971; Naftulin, Ware, and Donnelly 1983). In fact, even factors like class size (smaller is better) tend to impact instructor evaluations (Neath 1996).

It has been argued that part of the difficulty with SETs stems from the lack of a clear definition of effective teaching (Sherman and Backburn 1975). Because of this, it can be difficult for students to evaluate teaching effectiveness or for administrators to determine what students are actually evaluating. The fact that students often tend to form opinions of a class and an instructor early in the course and subsequent learning or classroom experiences do not tend to change these perceptions lends credibility to these theories (Ortinau and Bush 1987).

Equally disturbing is the lack of correlation that SETs have with teaching experience. Zabaleta (2007), for example, found that years of teaching experience or whether the instructor is a professional or a graduate teaching assistant is not associated with SETs. In fact, instructors with less than one year of teaching experience had better evaluations than those with more than 10 years of experience, despite the fact that graduate teaching assistants (GTAs) gave much lower grades than professionals. Although burnout has been offered as one potential reason for these results, the fact that the sample included 77 GTAs and 59 lecturers or professors with over four years of teaching cast doubt on this explanation. As pointed out by Dunkin (2002), there are distinct differences between experienced and novice teachers, yet, student evaluations do not reflect these. This would suggest that other things besides teaching are being evaluated in SETs. Zabaleta (2007) has postulated that student evaluations, instead of being grounded in quality of instruction, course content and objectives, and teaching expertise, are instead based on extraneous factors outside these important components of teaching. Therefore, many feel that SETs are not a valid measurement of teaching ability or performance.

Some opponents of SETs fear that they may actually constitute a threat to academic freedom and entice faculty to reduce academic standards (Martin 1998; McPherson and Jewell 2007; Watchel 1998). Gillmore and Greenwald (1999) has been one of many researchers who have suggested that evaluations can be 'bought' by grades, and are not actually related to learning (e.g. Clayson 2005; Weinberg, Fleisher, and Hashimoto 2007). Johnson's (2003) research supports this theory, explaining that students give instructors who grade more leniently higher evaluations than harder grading instructors. In other words, students reward easier grading instructors with higher evaluations. Others have found this relationship more complicated. Birnbaum (2000) found that even the belief that grades can impact evaluations modifies both faculty and student behaviors and Zabaleta (2007) found that higher grades have a weaker relationship with SETs than lower grades.

Several studies have explored how faculty utilize the information they gain through SETs. Yao and Grady (2005) reported that most faculty pay attention to SETs and use them to help gauge and assess their own teaching. While some instructors report finding SETs useful only in some concrete areas of course design (i.e. handouts, number of assignments; Spencer and Flyr 1992), other faculty are able to use SET feedback to make changes in their instruction, either in delivery or content. The ability to make substantial changes, however, is often complicated by other factors. Yet, faculty can only improve their teaching if they know how, and are motivated to do so (Centra 1993; Jacobs 1987).

Gender and student evaluations of teaching (SETs)

Although there has been extensive research conducted on SETs with special focus on reliability and validity and their use among faculty and administration, very little research has focused on faculty's emotional responses to SETs and in particular, the potential impact of gender.

Research pertaining to performance feedback suggests that gender can impact how people view and incorporate feedback, and women might be more receptive to the opinions of others. As Lungren and Rudawsky (1998) has argued, women's socialization process tends to emphasize taking into account the feelings and perceptions of others, while males often rely more on impersonal standards. Stated differently, it has been theorized that women have a more 'collectivist' or 'relational self' and men have more of an 'autonomous' or 'individualistic self'. A relational self is more attentive to others and accepting of their feedback. An autonomous self is more independent and resistant to the influence of others (Lungren and Rudawsky 1998).

When evaluating how feedback from others impacts self-assessment, it has been suggested that males and females are impacted similarly when the feedback is objective, but when the feedback is subjective, females evaluate themselves more negatively than males (Dweck and Bush 1976; Nicholls 1975). This line of research surmises that when women are given objective feedback, they do not think less of their own performance than males. But when the feedback is subjective, women's self-evaluations are more negative than males. It is hypothesized that this could lead to emotional differences in how male and female faculty members react to subjective evaluative feedback, such as the commonly used SETs.

Current study

To assess the perceptions of SETs, all teaching faculty in the College of Veterinary Medicine and Biomedical Sciences (CVMBS) at a western Land Grant University were asked to complete an anonymous online survey. This study was approved by the institution's Research Integrity and Compliance Review Office. The CVMBS is home to 219 faculty and four departments (Clinical Science; Biomedical Sciences; Environmental and Radiological Health Sciences; and Microbiology, Immunology, and Pathology). Many of these faculty members teach courses within their home department as well as courses within the Professional Veterinary Medicine (PVM) program.

The survey included demographic questions including gender, rank (assistant, associate, and full professor), and position (non-tenure track, tenure track, and tenured). The survey asked faculty to indicate their feelings and thoughts while reading their teaching evaluations from students. Examples of feelings/thoughts include 'frustrated', 'validated for the work I have put into teaching', and 'satisfied'. Other questions focused on aspects in addition to teaching that might impact students' evaluations (i.e. students' age and faculty members' gender). Additionally, agreement level (from strongly agree to strongly disagree) for several statements about SETs were included (e.g. 'Student ratings reflect the quality of instruction well' and 'Faculty members in general tend to water down their requirements in order to get favorable ratings'). Lastly, faculty members were asked to indicate how/if they use evaluations to alter either the material they teach, or the manner in which they teach (See the Appendix for survey).

In addition to obtaining the subjective perceptions of faculty regarding their SETs, objective data collected as part of the SETs were gathered to determine if SET results were statistically different based on gender or rank. All course evaluations for the PVM program share a common question pertaining to teaching effectiveness. The question asks students to rate their instructor on teaching effectiveness on a scale of 1 (very effective) to 5 (not effective). Other departments do not use the same question so this information was not available for all faculty of the college, yet because many faculty from each department teach within the PVM program, this information was available for a majority of the CVMBS faculty.

Results

All CVMBS teaching faculty were asked to complete the online anonymous survey, resulting in 75 completed surveys (from a total sample of 219; 34.2%). Included in the sample were 43 females (57.3%) and 32 males (42.7%). Rank level of respondents included: assistant professor 31 (41.3%), associate professor 20 (26.7%), full professor 23 (30.7%), and other 1 (1.3%). Position type included non-tenure track 16 (21.6%), tenure track 25 (33.8%), and tenured 33 (44.6%). This sample is representative of the faculty in the college.

Objective of student evaluations of teaching (SETs) analysis

This study was designed to determine if gender, rank, or position impact how faculty feel about their student evaluations. It was important, however, to determine if the

evaluations were actually different for these groups. To determine if student evaluations were significantly different based on gender or rank (data were not available for position), the general question related to teaching effectiveness was analyzed. This question asked students to rate their instructor on a scale from 1 (very effective) to 5 (not effective) with the following directions:

> Please use the following description of effective teaching as the criteria to rate this instructor: an effective instructor is one who comes to class well prepared, delivers content clearly and enthusiastically using a variety of presentation methods, and fosters an active learning environment by encouraging students to ask questions, join in discussions, and otherwise participate in learning activities.

No significant differences were found in these values based on instructor gender, rank, or position (see Table 1).

Survey results

A multivariate generalized linear model test was conducted to determine differences in responses based on gender, position, or rank. Significance level was accepted at $p < 0.05$. Main effects for each variable (i.e. gender, position, and rank), controlling for the other two variables, are reported.

Faculty were asked three questions pertaining to their use of evaluations. These included: 'I use the evaluations to alter what material I teach', 'I use the evaluations to alter how I teach my material', and 'I do not change my teaching based on student evaluations'. No differences were found based on rank or position. Females reported being less likely to change the material they taught ($F = 5.13$, $p = 0.027$) than males.

In addition to questions pertaining to the usage of evaluations, faculty were asked to indicate what factors they feel might influence student evaluations. No differences were found based on rank, and the only item found to be statistically different based on position was 'In general the evaluations do not provide much useful information', with a higher percentage of non tenure-track faculty agreeing to this statement compared to tenure-track or tenured faculty ($F = 3.30$, $p = 0.043$).

Table 1. Instructor gender, rank, and position.

	M	SD	F	p
Gender			1.62	0.21
Male ($n = 75$)	1.76	0.39		
Female ($n = 62$)	1.86	0.47		
Rank			2.3	0.06
Assistant ($n = 38$)	1.82	0.48		
Associate ($n = 33$)	1.73	0.33		
Full ($n = 38$)	1.73	0.44		
Student ($n = 7$)	2.18	0.63		
Other ($n = 20$)	1.90	0.31		
Position			2.27	0.07
Non tenure-track ($n = 37$)	1.96	0.47		
Tenure track ($n = 26$)	1.81	0.45		
Tenured ($n = 73$)	1.71	0.36		

Several items were found to be statistically different, however, for males and females. When asked 'How much impact do you think your gender has on their evaluation of you?', females reported a strong or moderate impact more often than males ($F = 14.23$, $p < 0.000$). Females indicated higher levels of agreement to the following statements 'In general the evaluations do not provide much useful information', 'Student ratings tend to be more of a popularity contest', 'One does not have to be a good instructor in order to get positive evaluations', and 'Student ratings undermine student–instructor relations'. Males indicated higher levels of agreement to the following statements: 'Student ratings reflect the quality of instruction well' and 'Student ratings help instructors improve their treatment of students'.

Feelings and thoughts related to evaluations

Faculty members were asked about the feelings and thoughts they had while reading their student evaluations. They were given a list of emotions and asked to indicate how well each of these reflected their own experiences while reading their teaching evaluations. Because many of these feelings are highly correlated, three groups of thoughts/feelings were created. The first cluster is labeled 'unhappy' and includes 'anxious', 'disheartened', 'depressed', and 'worried' (Cronbach's Alpha = 0.920). The second cluster is labeled 'mad' and includes: 'frustrated', 'angry', 'irritated', and 'disgusted' (Cronbach's Alpha = 0.875). The last cluster is labeled 'positive' and includes 'validated', 'energized', 'uplifted', 'motivated', 'satisfied', 'appreciative', and 'happy' (Cronbach's Alpha = 0.911). No differences based on rank or position were found for any of these three clusters. A gender difference was found, however, for 'unhappy' ($F = 4.30$, $p = 0.042$), and 'mad' ($F = 4.50$, $p = 0.038$). No gender difference was found for the cluster 'positive'.

Discussion

Our study focuses on how faculty members respond to student evaluations. While minimal differences were found in responses based on rank or position, several differences were found based on faculty gender. Overall, female faculty appear to be more negatively impacted by student evaluations than male faculty. Female faculty are more likely to feel emotions related to unhappiness and anger after reading SETs than male faculty. This difference is especially noteworthy due to the fact that female faculty's SETs are not statistically different than that of their male colleagues. These gender differences support previous research that suggests males and females receive and react differently to personal evaluation.

It is perhaps also not surprising, given the negative impact of SETs on female faculty, that they report being less likely to change the material they teach based on SETs. Similar to our study, Nasser and Fresko (2002) found more males reported modifying their teaching based on SETs than females. It appears clear that male and female faculty feel differently, and are therefore impacted differently, by their student evaluations. One explanation for this difference could be the difference in how boys and girls are socialized and the resultant long lasting impact on how males and females react to feedback. Boys tend to receive more evaluative attention in classrooms than girls, and more negative feedback in particular (Berk and Lewis 1977). Feedback for boys tends to focus more on non-intellectual qualities like

misconduct. Since the feedback is not directed toward their abilities, it has been suggested that boys can more easily attribute the feedback to a negative attitude of the teacher, rather than oneself (Dweck et al. 1978). The fact that negative feedback is used less frequently for girls and tends to focus on intellectual inadequacies makes it harder for girls to discount negative feedback (Roberts and Nolen-Hoeksema 1989). It has been suggested that these experiences shape adults, in that women, as compared to men, perceive evaluations as more informative about their abilities and are therefore more influenced by them.

Perhaps men place less value on negative evaluations by others and therefore are impacted less. Roberts and Nolen-Hoeksema (1989) found that men's self-perception is more likely to be influenced by positive rather than negative feedback. This difference was not found for women, however. They found that women tend to modify their self-image, based on the evaluations of others, more readily than men. This may be in part explained by the self-enhancing bias in men (a focus on internal attributes for successes) and the self-derogatory bias in women (a focus on external attributions for their success (Beyer 1990).

Women also incorporate both negative and positive feedback into their decision making, while men tend to incorporate more positive feedback and less negative feedback (Roberts 1991). Other studies have supported the idea that women perceive negative feedback as more negative than men and are therefore more highly impacted (Djamasbi and Loiacono 2008; Lungren and Rudawsky 1998; Rhodewalt and Hill 1995). Our results are consistent with Kling et al. (1999) who explained that because men's self-esteem is derived more from achievement of goals while women's self-esteem is more closely linked to relationships with others, that men and women will feel differently about perceived rejection or negative evaluations. They state that 'rejection will provide a more direct challenge to women's self esteem...For men, however, the rejection would not challenge their self esteem so directly, which may provide them with a coping advantage...' (491).

Baldwin, Granzberg, and Pritchard's (2003) research supports this premise, finding that rejection leads women to be more self-critical, while men respond to rejection more defensively and in a more compensatory manner. As Roberts and Nolen-Hoeksema (1994) pointed out, women tend to be more responsive to all evaluations, good and bad, than men. The literature would seem to support the fact that men and women differ in how they interpret and utilize feedback (Djamasbi and Loiacono 2008). This is supported by the results of the current study.

This is not to suggest that SETs cannot play an important role in evaluating faculty and giving students a voice. If SETs are constructed in such a way as to foster constructive feedback, they can be a useful tool. If feedback is unconstructive, however, it can be argued that it might be more damaging for female faculty than male faculty. It would appear, therefore, that designing SETs to offer the best opportunity for constructive feedback is of paramount importance.

Conclusions and recommendations

Heated debate on the usefulness of SETs in evaluating faculty teaching continues. Student feedback can be a useful tool, but one that carries both benefits and costs. The benefits to faculty, departments, and students include the potential to improve teaching, reduced uncertainty pertaining to student perceptions, and

increased self-esteem and confidence in teaching among faculty. The costs when faced with negative, unconstructive feedback, however, can include faculty's loss of confidence and reduced self-image (Miller and Karakowsky 2005).

There have been several studies arguing that SETs should not be used for career decisions, such as tenure, promotion, or raises, because SETs do not actually evaluate teaching. Indeed, it has also been suggested that one way to reduce grade inflation might be to limit the power of SETs (Zabaleta 2007). Other avenues to improve teaching effectiveness that have been suggested include teaching portfolios, student interviews, class observations and unbiased peer evaluations, and self-evaluations (Wright 2006; Zabaleta 2007).

Created carefully, SETs designed to foster constructive criticism are still viewed by many as potentially useful tools to improve teaching. Brightman (2005), for example, suggests that SETs should be normed. This would entail taking into consideration numerous factors (i.e. type of class, size, and time in which it is offered). Additionally, making a change from anonymous to confidential surveys has been suggested. As Wright (2006) pointed out, anonymous SETs were created with positive intent. The basic idea was to give students the ability to express themselves without fear of reprisal from faculty. Yet, with anonymous surveys, come potential problems. When anonymous, Wright (2006) suggests that students take no responsibility for their evaluations. It also eliminates the ability to follow-up on the results. For example, there is no way to determine if students who gave poor evaluations were present for most of the class periods or were performing well in the class. It is also not possible to track whether this person is a chronic complainer, or if this evaluation differs substantially from their other SETs. One solution is the creation of confidential, but not anonymous surveys. This would entail follow-up with individual students to enhance communication between students and administration, as well as the opportunity to conduct relevant research (e.g. correlations between evaluations and actual grades/performance).

Additionally, given the results of this study, it is suggested that information about the potential negative impact of SETs for faculty, and female faculty in particular, should be shared with faculty and administration. Supporting faculty in the process of evaluation might include educating them to the fact that the feelings they have about SETs are shared by many other faculty. It might also be useful to offer training to faculty on best to interpret their SETs, thereby promoting positive changes in teaching (Lang and Kersting 2007). Additionally, Brightman's (2005) study explores the importance of a mentoring program and provides some specifics on how to create a successful mentoring program for faculty. Mentoring can perhaps not only help faculty avoid burnout, but could also be used to minimize damage done by SETs.

Limitations to the current study include the fact that faculty in only one college were surveyed. Clearly the need for additional research involving larger, multi-institutional samples are needed to better understand the emotional impact of SETs on faculty. Since so many institutions of higher learning include SETs in their evaluation process, it is crucial to examine the impact of SETs on faculty.

It is critical that attention to the potential detrimental impact of these evaluations on faculty be made a top priority. SETs offer one avenue to assess teaching effectiveness, but the current status of SETs in many universities (anonymous with no support or training for faculty in how to interpret the results) creates a fertile ground

for negative outcomes. Small changes, including the change from anonymous to confidential, as well as professional development training for faculty might help make SETs ultimately more useful with less potential for damage.

References

Aleamoni, L.M. 1999. Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education* 13: 153–66.

Ambady, N., and R. Rosenthal. 1993. Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology* 64: 431–41.

Baird, J.S. 1987. Perceived learning in relation to student evaluation of university instruction. *Journal of Educational Psychology* 79: 90–1.

Baldwin, M.W., A. Granzberg, and E.T. Pritchard. 2003. Cued activation of relational schemas: Self-evaluation and gender effects. *Canadian Journal of Behavioural Science* 35: 153–63.

Berk, L.E., and N.G. Lewis. 1977. Sex role and social behavior in four school environments. *Elementary School Journal* 77: 205–17.

Beyer, S. 1990. Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology* 59: 960–70.

Birnbaum, M.H. 2000. A survey of faculty opinions concerning student evaluation of teaching. http://www.psych.fullerton.edu/mbirnbaum/faculty3.htm (accessed May 3, 2010).

Brightman, H.J. 2005. Mentoring faculty to improve teaching and student learning. *Decision Sciences Journal of Innovative Education* 3: 191–203.

Centra, J.A. 1993. *Reflective faculty evaluation*. San Francisco, CA: Jossey-Bass.

Clayson, D.E. 2005. Performance overconfidence: Metacognitive effects or misplaced student experience. *Journal of Marketing Education* 27: 122–9.

Clayson, D.E. 2009. Student evaluations of teaching: Are they related to what students learn? *Journal of Marketing Education* 31: 16–30.

Clayson, D.E., and D.A. Haley. 1990. Student evaluations in marketing: What is actually being measured? *Journal of Marketing Education* 12: 9–16.

Cohen, P.A. 1981. Student ratings of instruction and student achievement: A meta analysis of multi-section validity studies. *Review of Educational Research* 51: 281–309.

Comm, C.L., and D.F.X. Manthaisel. 1998. Evaluating teaching effectiveness in America's business schools: Implications for service marketers. *Journal of Professional Service Marketing* 16: 163–70.

Costin, E., W.T. Greenough, and R.J. Menges. 1971. Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research* 41: 511–35.

Djamasbi, S., and E.T. Loiacono. 2008. Do men and women use feedback provided by their Decision Support Systems (DSS) differently? *Decision Supports Systems* 44: 854–69.

Dowell, D.A., and J.A. Neal. 1982. A selective review of the validity of student ratings of teaching. *Journal of Higher Education* 53: 51–62.

Dowell, D.A., and J.A. Neal. 1983. The validity and accuracy of student ratings of instruction: A reply to Peter A. Cohen. *Journal of Higher Education* 54: 459–63.

Dunkin, M.J. 2002. Novice and award winning teachers' concepts and beliefs about teaching in higher education. In *Teacher thinking, beliefs, and knowledge in higher education*. ed. N.G. Hativa, 41–58. Dordrecht: Kluwer Academic.

Dweck, C.S., and E.S. Bush. 1976. Sex differences in learned helplessness: I. Differential debilitation with peer and adult evaluators. *Developmental Psychology* 12: 147–56.

Dweck, C.S., W. Davidson, S. Nelson, and B. Enna. 1978. Sex differences in learned helplessness: II the contingencies of evaluative feedback in the classroom and II. An experimental analysis. *Developmental Psychology* 14: 268–76.

Gillmore, G.M., and A.G. Greenwald. 1999. Using statistical adjustment to reduce biases in student ratings. *American Psychologist* 54: 518–9.

Hamermesh, D., and A. Parker. 2003. *Beauty in the classroom: Professor's pulchritude and putative pedagogical productivity*. Cambridge, MA: National Bureau of Economic Research.

Jacobs, L.C. 1987. *University faculty and students' opinions of student ratings.* Bloomington, IN: Bureau of Evaluative Studies and Testing.

Johnson, V.E. 2003. *Grade inflation: A crisis in college education.* New York: Springer.

Kling, K.C., J.S. Hyde, C.J. Showers, and N.S. Buswell. 1999. Gender differences in self esteem: A meta-analysis. *Psychological Bulletin* 125: 470–500.

Lang, J.W.B., and M. Kersting. 2007. Regular feedback from student ratings on instruction: Do college teachers improve their ratings in the long run? *Instructional Science* 35: 187–205.

Lungren, D.C., and D.J. Rudawsky. 1998. Female and male college students' responses to negative feedback from parents and peers. *Sex Roles* 39: 409–29.

Marsh, H.W. 1977. The validity of student's evaluations: Classroom evaluations of instructors independently nominated as best and worst teachers by graduating seniors. *American Educational Research Journal* 14: 441–7.

Marsh, H.W. 1984. Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and utility. *Journal of Educational Psychology* 76: 707–54.

Marsh, H.W. 1987. Students' evaluations of university teaching: Research findings, methodological issues and directions for future research. *International Journal of Educational Research* 11: 253–388.

Martin, J. 1998. Evaluating faculty based on student opinions, problems, implications and recommendations from Deming's theory of management perspective. *Issues in Accounting Education* 13: 1079–94.

McKeachie, W.J. 1986. *Teaching tips.* Lexington, MA: D.C. Health.

McPherson, M.A., and R.T. Jewell. 2007. Leveling the playing field: Should student evaluation scores be adjusted? *Social Science Quarterly* 88: 868–81.

Miller, D.L., and L. Karakowsky. 2005. Gender influences as an impediment to knowledge sharing: When men and women fail to seek peer feedback. *The Journal of Psychology* 139: 101–18.

Naftulin, D.H., J.E. Ware, and E.A. Donnelly. 1983. The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education* 75: 138–49.

Nasser, F., and B. Fresko. 2002. Faculty views of student evaluation of college teaching. *Assessment and Evaluation in Higher Education* 27: 187–98.

Neath, I. 1996. How to improve your teaching evaluations without improving your teaching. *Psychological Reports* 78: 1363–72.

Nicholls, J.G. 1975. Causal attribution and other achievement-related cognitions: Effects of task outcome, attainment value and sex. *Journal of Personality and Social Psychology* 31: 379–89.

Olshavsky, R.W., and R.A. Spreng. 1995. Consumer satisfaction and students: Some pitfalls of being customer driven. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior* 8: 69–77.

Ortinau, D.J., and R.P. Bush. 1987. The propensity of college students to modify course expectations and its impact on course performance information. *Journal of Marketing Education* 9: 42–52.

Ramsden, P. 2003. *Learning to teach in higher education.* London: Taylor & Francis.

Rhodewalt, F., and S.K. Hill. 1995. Self handicapping in the classroom – the effects of claimed self handicaps on responses to academic-failure. *Basic and Applied Social Psychology* 16: 397–416.

Roberts, T. 1991. Gender and the influence of evaluations on self-assessments in achievement settings. *Psychological Bulletin* 109: 297–308.

Roberts, T., and S. Nolen-Hoeksema. 1989. Sex differences in reactions to evaluative feedback. *Sex Roles* 21: 725–47.

Roberts, T., and S. Nolen-Hoeksema. 1994. Gender comparisons in responsiveness to others' evaluations in achievement settings. *Psychology of Women Quarterly* 18: 221–40.

Rodin, M., and B. Rodin. 1972. Student evaluations of teachers. *Science* 177: 1164–6.

Sebastian, R.J., and D. Bristow. 2008. Formal or informal? The impact of style of dress and forms of address on business students' perceptions of professors. *Journal of Education for Business* 83: 196–201.

Sherman, B.R., and R.T. Backburn. 1975. Personal characteristics and teaching effectiveness of college faculty. *Journal of Educational Psychology* 67: 124–31.

Spencer, P.A., and M.L. Flyr. 1992. *The formal evaluation as an impetus to classroom change: Myth or reality?* Riverside, CA: University of California Press.

Sproule, R. 2000. Student evaluation of teaching: A methodological critique of evaluation practices. *Education Policy Analysis Archives.* http://www.epaa.asu.edu/epaa/v8n50.html (accessed September 2, 2009).

United States Department of Education. 2006. *A test of leadership: Charting the future of U.S. higher education.* Jessup, MD: United States Department of Education.

Watchel, K.H. 1998. Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education* 23: 191–211.

Weinberg, B.A., B.M. Fleisher, and M. Hashimoto. 2007. Evaluating methods for evaluating instruction: The case of higher education (NBER Working Paper No. 12844). *National Bureau of Economic Research.* http://www.nber.org/papers/w12844 (accessed September 2, 2009).

Weinberg, B.A., M. Hashimoto, and B.M. Fleisher. 2009. Evaluating teaching in higher education. *Journal of Economic Education* 40: 227–54.

Wilson, K.L., A. Lizzo, and P. Ramsden. 1997. The development, validation, and application of the Course Experience Questionnaire. *Studies in Higher Education* 22: 33–54.

Wright, R.E. 2006. Student evaluations of faculty: Concerns raised in the literature, and possible solutions. *College Student Journal* 40: 417–22.

Yao, Y., and M.L. Grady. 2005. How do faculty make formative use of student evaluation feedback?: A multiple case study. *Journal of Personnel Evaluation in Education* 18: 107–26.

Yunker, P.J., and J.A. Yunker. 2003. Are student evaluations of teaching valid? Evidence from an analytical business core course. *Journal of Education for Business* 78: 313–7.

Zabaleta, F. 2007. The use and misuse of student evaluations of teaching. *Teaching in Higher Education* 12: 55–76.

Appendix. Faculty survey about Professional Veterinary Medicine (PVM) didactic course/instructor evaluations by students

For classification purposes only, please answer the following:

Your gender: male   female

Your rank: assistant professor   associate professor   professor
                tenuredtenure-track   special appointment

Number of years of college teaching experience _____.

	Strongly agree	Agree	Neutral	disagree	Strongly disagree
I use the evaluations to alter *what material* I teach					
I use the evaluations to alter *how I teach* my material					
I do not change my teaching based on student evaluations					
I think student evaluations could be made more useful by making them confidential, rather than anonymous (whereby instructors do not have access to students' names but some designated person does)					

If you do change your teaching based on evaluations,

Do you:

        Make changes based on one semester of feedback.

        Make changes based on 2 semesters of feedback.

        Make changes based on longer term trends (more than 2 semesters).

Please indicate how well each of the following reflects your feelings/thoughts when you read your teaching evaluations from students:

	Strongly reflect	Moderately reflect	Minimally reflect	Don't reflect at all
Motivated to improve my teaching				
Frustrated				
Anxious				
Disheartened				
Uplifted				
Depressed				
Angry				
Validated for the work I have put into teaching				
Amused				
Satisfied				
Appreciative				
Irritated				
Happy				
Relieved				
Disgusted				
Energized				
Worried				

Comments:

Please indicate the impact level you feel the following aspects have on student's evaluations:

	Strong impact	Moderate impact	Minimal impact	No impact
If a student expects to receive a high grade in your class, how much impact do you feel this has on their evaluation of you?				
How much impact do you think *students* age/experience has on their evaluation of you				
How much impact do you think *your* age/experience has on their evaluation of you				
How much impact do you think your gender has on their evaluation of you				

Please indicate how much you agree or disagree with the following statements:

	Strongly agree	Agree	Neutral	disagree	Strongly disagree
Most students take the evaluation process seriously					
In general the evaluations do not provide any useful information					
Faculty members in general tend to water down their requirements in order to get favorable ratings					
Administering the course and teacher eva luations every semester is a waste of time					
Instructors who demand a lot from their students get low evaluations					
Student ratings measure how nice an in structor is					
Student ratings reflect the quality of instruction well					
One does not have to be a good instructor in order to get positive evaluations					
Student ratings are an acceptable criterion for granting tenure					
Student ratings are an acceptable criterion for granting promotion					
Student ratings are an acceptable criterion for annual evaluation					
Student ratings are a good basis for planning changes in teaching					
Student ratings undermine student–instructor relations					
Student ratings help instructors improve their treatment of students					