REVIEW OF THE VARIABLES THAT INFLUENCE STUDENT EVALUATION OF TEACHING

Prepared for the UBC Student Evaluation of Teaching Committee
February 25, 2013

TABLE OF CONTENTS

SUMMARY

The purpose of this review was to examine the variables that influence student evaluation of teaching, to summarize the research evidence, as well as to report current practices at other institutions. Variables that influence ratings can be grouped into characteristics of the student, instructor, or course. Evidence from published research is inconsistent and there are both confirmatory and contradictory findings. Results from our review, however, suggest that the most consistent findings are the effects of student and course grades, as well as field of study. We also show that these effects are small and a large proportion of the variability in teaching evaluations remains unknown. Finally, an informal poll from institutions in the United States showed that most do not compare instructors at the university-level, but some include student, instructor, and/or course variables in their evaluations. We conclude that there are potential biases in teaching evaluations, but the amount of this bias is small.

TYPES OF VARIABLES

There have been numerous attempts to examine the variables that influence instructor ratings, and it continues to be an area of research (recent reviews: Marsh, 2007; Pounder, 2007). There are three groups of variables or characteristics that may influence instructor ratings:

1. Student variables (e.g., student gender, grade in course, class attendance)
2. Instructor variables (e.g., instructor gender, age, rank, personality)
3. Course variables (e.g., elective, year-level, class size, field of study, average course grade)

This review will examine research evidence for the variables that influence instructor, but not course, ratings. As The University of British Columbia (UBC) examines instructor effectiveness through the six University Module Items, variables that influence course ratings will not be examined in this review.

RESEARCH EVIDENCE

Student Variables

Student Gender

The influence of a student's gender on instructor ratings is a controversial issue and research thus far has been inconsistent. There may be an interaction between student and instructor gender; that is, female students may rate male and female instructors differently than do their male counterparts. Specifically, female students rated female instructors more positively than male instructors, but male students rated both genders equally (Bachen et al., 1999; Hakstian, 2009). However, other research has found that female students rated all instructors higher than male students (Fernandez et al., 1997; Norvilitis et al., 2009), both female and male students had same-gender preferences (Centra et al., 2000), or there was no effect of student gender on ratings (Clayson, 1999; Davidovitch et al., 2006; Heckert et al., 2006; Marsh et al., 1997). These results suggest that while male and female students may rate their instructors differently, these differences are small and do not consistently impact evaluations.

Student Grade

A large body of research has focused on the influence of students' grades on teaching evaluations, whereby students with higher grades may give more positive evaluations to their instructors than those with lower grades (Pounder, 2007; Marsh, 2007). Meta-analyses have shown that student grades account for 10% to 14% of the variance in teaching evaluations (Cohen, 1983; McCallum, 1984).

Furthermore, this relationship is specific to a student's grade in a course and not their cumulative grade or grade point average (Brockx et al., 2011; Langbein, 1994).

Other research has shown that expected grades also influence teaching evaluations (Beran et al., 2005; Heckert et al., 2006; Isley et al., 2005; Johnson et al., 1990; Langbein, 1994; McPherson, 2006; McPherson et al., 2007, 2009; Perkins et al., 1990). These results imply that students evaluate an instructor's performance relatively to how they believe they will perform in the class. However, the expected grade effect is eliminated when a student's prior subject interest is accounted for (Marsh, 1983, 1987) or when perceived learning is controlled for (Centra, 2003; Marsh et al., 2000). Finally, it has been shown that expected grades might influence course ratings, but not instructor ratings (Patrick, 2011).

Therefore, these results suggest that a student's grade may influence teaching evaluations, but there is a large proportion of variance that is unaccounted for (i.e., 86% to 90%). These findings also suggest that the effect of expected grades on instructor ratings are contaminated by other factors such as the student's interest in the subject and perceived learning.

Class Attendance

Similarly to higher student grades, students with higher attendance records rate their instructors more positively (Arnold, 2009; Beran et al., 2005; Brockx et al., 2011; Davidovitch et al., 2006). For example, students who attended less than 40% of classes rated their instructors lower than those who attended more than 40% of classes (Davidovitch et al., 2006). Brockx et al. (2011) showed that class attendance predicted 5% of variance in instructor ratings, which is a significant, but small amount. Therefore, these results show that higher class attendance is associated with better teaching evaluations, but this is a small effect.

Instructor Variables

Instructor Gender

The influence of instructor gender on ratings is based on the assumption that female instructors may be discriminated against in higher education; however, the research evidence for this premise is lacking. In general, when gender differences are found, more positive evaluations are shown for female instructors (Feldman, 1993; Hakstian, 2009; Heckert et al., 2006; Marsh et al., 1997). There have been some contradictory findings, however, where male instructors were rated higher than female instructors (Kierstead et al., 1988; McPherson et al., 2009) or where there was no effect of instructor gender (Daud et al., 2011; Koh et al., 1997; McPherson et al., 2007). Also, any differences between genders may depend on the evaluation question (e.g., teacher's sensitivity). Overall, the effect of instructor gender on ratings is minimal, with less than 1% of variance accounted for in instructor evaluations (Feldman, 1993).

Instructor Age, Rank, and Experience

Intuitively, instructors with more experience should have higher evaluation ratings as they have had greater opportunities to improve their teaching. However, the effect of instructor age, rank, and experience on teaching evaluations has been inconsistent.

Most studies do not show an effect of instructor age on ratings (Daud et al., 2011; Feldman, 1983; Heckert et al., 2006; Koh et al., 1997). Others have shown a negative relationship between age and ratings (Clayson, 1999) or a curvilinear relationship where younger and older instructors obtain higher ratings than middle-aged instructors (Kinney et al., 1992). However, Kinney et al. (1992) also showed

that the age effect differs between academic disciplines and it is minimal in comparison to other factors.

In contrast to expectations, several studies have shown that higher-ranked instructors with more experience (i.e., tenured faculty) have lower ratings than their less-experienced colleagues (Blackhart et al., 2006; Clayson, 1999; McPherson et al., 2009). Other research has suggested that ratings increase with experience up to a point, but then they decline (Langbein, 1994). Finally, there may be no relationship between experience and instructor ratings (Daud et al., 2011; Feldman, 1983; Koh et al., 1997; Marsh et al., 1991). Overall, these results show that the influence of age, rank, and experience on teaching evaluations is minimal.

Instructor Personality

One area of research that has gained attention in teaching evaluations is the effect of an instructor's personality. Traits such as extraversion, expressiveness, leadership, dominance, and warmth have shown a significant relationship with student's ratings (Chen et al., 2010). In psychology, there are the "Big Five" factors of personality, including extraversion, neuroticism, agreeableness, conscientiousness, and openness; several studies have shown that all factors except for neuroticism lead to higher instructor ratings (Chen et al., 2010; Kniepp et al., 2010; Patrick, 2011; Williams et al., 1997).

However, the effect of instructor personality on ratings has varied dramatically between studies; for instance, Feldman (1986) showed minimal effects of conscientiousness and neuroticism, Chen et al. (2010) showed that 2.6% of the variance in ratings was due to personality, and Clayson (1999) showed that up to 80% of ratings was attributable to personality. The differences in findings are likely due to how personality was defined and how teaching effectiveness was measured. Therefore, although a teacher's personality may influence their ratings, further research evidence is needed.

Course Variables

Elective Course

Students who are taking a course as an elective may rate their instructors more highly than those who are required to take the course (Centra, 2003; Davidovitch et al., 2006). However, some studies have found an interaction between required courses and class size, which may lead to the lower ratings in required courses (Pounder, 2007), or that the effect of required courses on ratings differs between social sciences, humanities, and biological sciences (Kinney et al., 1992). Therefore, these results suggest that instructors of elective courses may receive more favourable ratings than those who teach required courses, but the effect of this relationship may be confounded by other factors.

Year-Level of Course

The year-level of a course (i.e., first year vs. fourth year) may impact instructor ratings due to several factors, such as class size, student interest, and the amount of teacher-student interactions. For example, it has been shown that higher-level courses have more positive ratings than lower-level courses (Brown, 1976; Frey et al., 1975; Laws et al., 2010). Similarly to other variables, the effect of year-level differs between fields of study or academic disciplines (Kinney et al., 1992). Finally, some researchers have shown that introductory-level courses have more positive ratings than higher-level courses, depending on the question (Buchert et al., 2008), or that there is no effect of year-level on instructor evaluations (Centra, 2003). These results suggest that the year-level of a course may have a potential impact on ratings, but this relationship is inconsistent.

Class Size

It is commonly assumed that smaller classes have higher instructor ratings, as there are greater opportunities for teacher-student interactions. Many studies have shown this effect (Bedard et al., 2008; Centra, 2003; Feldman, 1984; Hakstian, 2009; Isley et al., 2005; Koh et al., 1997; Liaw et al., 2003; McPherson, 2006; Toby, 1993).

In a meta-analytic review, Feldman (1984) showed that less than 1% of variance in teaching evaluations was attributable to class size. In contrast, Bedard and Kuhn (2008) showed that up to 29% of the variability in ratings was due to class size. These studies also showed that linear relationships might not reflect the actual relationship between class size and ratings. It may in fact be curvilinear, such that small and large classes have higher ratings than middle-sized classes (Feldman, 1984) or it may be cubic, where the difference in ratings is more drastic for small- to medium-sized classes than for large courses (Bedard et al., 2008). Finally, a few studies have shown no relationship between class size and instructor ratings (Krautmann et al., 1999; Langbein, 1994). Overall, these results illustrate a potential relationship between class size and ratings; however, the size of this effect varies between studies.

Course Grades

The contentious issue of course grades is whether instructors can "buy" better ratings by giving higher grades. However, others suggest that higher course grades are a reflection of teaching effectiveness (Brockx et al., 2011; Marsh, 2007). To this end, much research has focused on the effect of course grades on teaching evaluations.

A positive relationship between average course grades and instructor ratings has been shown, with 5% to 13% of the variance in ratings being accounted for by grades (e.g., Arnold, 2009; Blackhart et al., 2006; Brown, 1976; DeBerg et al., 1990; Hakstian, 2009). For example, a 1-point increase in grade point average leads to improvement in instructor rating (Brockx et al., 2011; Krautmann et al., 1999). Most of these improvements are small, with a 0.34 to 0.56 point increase on a 5-point scale (Krautmann et al., 1999). Other research, however, has shown no effect of course grades on teaching evaluations (Langbein, 1994).

These results suggest that course grades are a potential predictor of instructor ratings, but this effect is small. Similarly to other factors, a large proportion of the variability in evaluations remains unaccounted for when only one variable is measured.

Field of Study

The field of study or academic discipline, such as faculty, division, school, or department, may have an effect on instructor ratings. In general, higher instructor ratings are shown in the arts and humanities, followed by biological and social sciences, business, computer science, mathematics, engineering and physical sciences (Beran et al., 2005; Cashin, 1990; Davidovitch et al., 2006; Fernandez et al., 1997; Ory, 2001). Furthermore, several studies have shown that the effects of instructor age, class size, year-level of the course, and required courses differs between fields of study (Centra, 2003; Cranton et al., 1986; Kinney et al., 1992), which implies that academic discipline influences instructor ratings. These results show that field of study may influence teaching evaluations, but there is likely an interaction between the various student, instructor, or course variables.

Combination of Variables

The analysis of multiple variables is more likely to capture the predictors of teaching effectiveness. However, the types of variables included in published studies have varied and thus, it remains difficult to interpret across them. As shown in Table 1, the number of variables in these studies ranged from 2 to 42, the number of significant variables ranged from 2 to 26, and the proportion of variance accounted for in teaching evaluations (R^2) ranged from 7% to 78%. This table also shows that most variables predicted a small amount of variance in instructor ratings (i.e., less than 20%), which suggests that there are other predictors of instructor effectiveness that are not portrayed by student, instructor, or course variables.

Table 1. Studies that examined a combination of variables

Reference	Number of variables	Number of significant variables	R^2
Beran et al. (2005)	10	9	0.07
Arnold (2009)	6	3	0.08
Langbein (1994)	14	8	0.10
Blackhart et al. (2006)	2	2	0.12
Koh et al. (1997)	14	5	0.15
Chen et al. (2010)	9	2	0.17
Patrick (2011)	7	3	0.20
Isley et al. (2005)	16	3	0.22
Centra (2003)	42	26	0.72
Bedard et al. (2008)	8	8	0.77
McPherson (2006)	8	4	0.78

CURRENT PRACTICE AT OTHER INSTITUTIONS

An informal poll was conducted by email to determine the types of variables that are currently being used at other institutions. Specifically, respondents were asked whether their institution conducts university-level comparisons and what types of variables are being used in their current practices. There were a total of 34 respondents from across the United States.

Student Variables

Kansas State University includes two student variables in their evaluations, called ***student motivation*** and ***student effort.*** Specifically, these variables are measured by three questions on the evaluation form: "Your interest in taking this course before you enrolled," "Your effort to learning this course," and "The amount you have learned in this course." No other institutions reported that they use student variables for conducting university-level comparisons.

Instructor Variables

Two institutions indicated that they compare evaluations by ***instructor rank*** (e.g., tenured faculty, tenure-track faculty, non-tenured faculty, or teaching assistant). These institutions included: University of Colorado Boulder and University of Illinois Urbana-Champaign.

Course Variables

For institutions that conduct university-level comparisons, the most common comparison is by characteristics of the course. One institution indicated that they compare ***elective courses*** to required or mixed courses (University of Illinois Urbana-Champaign), one institution compares evaluations by ***credit hour*** of the course (University of Arizona), and three institutions compare instructors based on ***class size*** (Georgia Institute of Technology, Kansas State University, University of Arizona,). For example, Georgia Institute of Technology includes groups for <16, 16-35, 36-99, and >99 students in the course. Furthermore, six institutions compare their instructors based on ***year-level of the course***. These comparisons were either by each year-level separately (100, 200, 300, 400, etc.) or by broader levels of undergraduate versus graduate courses. These institutions included: American University, Florida State University (http://cat.fsu.edu/courseevaluations/tools/index.cfm), University of Arizona, University of Colorado Boulder, University of Maryland, and University of Notre Dame.

At the university-level, most institutions compare instructors using ***field of study.*** There were a total of nine institutions that compare their instructors at the university/campus, college/school/division, and department/program level, which included: American University, Boston College, Brigham Young University, Tulane University, University of California Riverside, University of Colorado Boulder, University of Illinois Urbana-Champaign, University of Notre Dame, and University of Texas at Austin. As discussed further below, we did not include in this section institutions that compare their instructors at the school- or department-level only (i.e., the specific school or department conducts their own comparisons). These results suggest that of the institutions that *do* compare their instructors at the university-level, the most common comparison is by field of study.

No Comparisons At The University-Level

The majority of respondents (20 out of 34, or 59%) indicated that they do not compare instructors at the university-level. These institutions included: Cornell University, Harvard University, Pennsylvania State University, Purdue University, St. Louis University, Syracuse University, University at Buffalo SUNY, University of California Irvine, University of California Los Angeles, University of California San Francisco, University of Dayton, University of Florida, University of Kansas, University of Kentucky, University of Miami, University of Pennsylvania, University of Pittsburgh, University of South California, University of Virginia, and Virginia Polytechnic Institute and State University.

The reasons for not comparing instructors at the university-level was due to several factors:

1. Individual colleges, schools, or departments conduct their own evaluations;
2. Some institutions do not have the technological capacity to compare instructors at this level;
3. Several indicated that university-level comparisons induce bias in evaluations and thus, they do not promote these types of comparisons.

Therefore, the current practice at most institutions is to ***not compare instructors at the university-level.*** However, this was an informal survey conducted through email and although we had a large response rate, many institutions or regions were underrepresented. These results should be interpreted with caution.

CONCLUSIONS

Our results show that the most consistent influences on the student evaluation of teaching are student grades, average course grades, and the field of study; however, these influences are small. These results suggest that student grades may bias instructor ratings, but it may also be a measure of teaching effectiveness. Furthermore, we showed that while evaluations may vary between fields of study or academic discipline, interaction effects might contaminate these results.

These results are similar to Marsh (2007) who suggested that there is little support for any variables that bias instructor ratings. When there are biases, the compromise is to construct normative scores for similar courses (e.g., undergraduate vs. graduate courses), but the research evidence for constructing norms is lacking. Finally, some researchers suggest to not control for biasing factors (e.g., Abrami et al., 1990), as the negative consequences may outweigh the positive (McKeachie, 1996).

In conclusion, this review shows that there are potential biasing factors in teaching evaluations, which may warrant the use of norms, but the amount of bias in instructor ratings is small.

REFERENCES

Abrami, P. C., d'Apollonia, S., & Cohen, P. A. (1990). Validity of student ratings of instruction: What we know and what we do not. *Journal of Educational Psychology, 82,* 219-231.

Arnold, I. J. M. (2009). Do examinations influence student evaluations? *International Journal of Educational Research, 48,* 215-224.

Bachen, C. M., McLoughlin, M. M., & Garcia, S. S. (1999). Assessing the role of gender in college students' evaluations of faculty. *Communication Education, 48,* 194-210.

Bedard, K., & Kuhn, P. (2008). Where class size really matters: Class size and student ratings of instructor effectiveness. *Economics of Education Review, 27,* 253-265.

Beran, T., & Violato, C. (2005). Ratings of university teacher instruction: How much do student and course characteristics really matter? *Assessment & Evaluation in Higher Education, 30,* 593-601.

Blackhart, G. C., Peruche, B. M., DeWall C. N., & Joiner, T. E. (2006). Factors influencing teaching evaluations in higher education. *Teaching of Psychology, 33,* 37-39.

Brockx, B., Spooren, P., & Mortelmans, D. (2011). Taking the grading leniency story to the edge. The influence of student, teacher, and course characteristics on student evaluations of teaching in higher education. *Educational Assessment, Evaluation and Accountability, 23,* 289-306.

Brown, D. L. (1976). Faculty ratings and student grades: A university-wide multiple regression analysis. *Journal of Educational Psychology, 68,* 573-578.

Buchert, S., Laws, E. L., Apperson, J. M., & Bregman, N. J. (2008). First impressions and professor reputation: Influence on student evaluations of instruction. *Social Psychology of Education, 11,* 397-408.

Cashin, W. E. (1990). Students do rate different academic fields differently. *New Directions for Teaching and Learning, 43,* 113-121.

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education, 44,* 495-518.

Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education, 71,* 17-33.

Chen, G.-H., & Watkins, D. (2010). Stability and correlates of student evaluations of teaching at a Chinese university. *Assessment & Evaluation in Higher Education, 35,* 675-685.

Clayson, D. E. (1999). Students' evaluation of teaching effectiveness: Some implications of stability. *Journal of Marketing Education, 21,* 68-75.

Cohen, P. A. (1983). Comment on a selective review of the validity of student ratings of teaching. *Journal of Higher Education, 54,* 448-458.

Cranton, P. A., & Smith, R. A. (1986). A new look at the effect of course characteristics on student ratings of instruction. *American Educational Research Journal, 23,* 117-128.

Daud, N. M., & Kassim, N. L. A. (2011). Examining student rating of teaching effectiveness using FACETS. *Journal of Applied Measurement, 12,* 135-143.

Davidovitch, N., & Soen, D. (2006). Class attendance and students' evaluation of their college instructors. *College Student Journal, 40,* 691-703.

DeBerg, C. L., & Wilson, J. R. (1990). An empirical investigation of the potential confounding variables in student evaluation of teaching. *Journal of Accounting Education, 8,* 37-62.

Feldman, K. A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education, 18,* 3-124.

Feldman, K. A. (1984). Class size and students' evaluations of college teacher and courses: A closer look. *Research in Higher Education, 21,* 45-116.

Feldman, K. A. (1986). The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: A review and synthesis. *Research in Higher Education, 24,* 139-213.

Feldman, K. A. (1993). College students' view of male and female college teachers: Part II – Evidence from students' evaluations of their classroom teachers. *Research in Higher Education, 34,* 151-211.

Fernandez, J., & Mateo, M. A. (1997). Student and faculty gender in ratings of university teaching quality. *Sex Roles, 37,* 997-1003.

Frey, P. W., Leonard, D. W., & Beatty, W. W. (1975). Student ratings of instruction: Validation research. *American Educational Research Journal, 12,* 435-444.

Hakstian, R. (2009). An investigation into the effects of instructor gender, field of study, and student-respondent gender on UMI scores in the 2008-09 SEoT administration. The University of British Columbia, obtained from http://teacheval.ubc.ca/files/2009/03/SEoT-Gender-X-Field-of-Study-Analysis-Revised-10-27-09.pdf

Heckert, T. M., Latier, A., Ringwald, A., & Silvey, B. (2006). Relation of course, instructor, and student characteristics to dimensions of student ratings of teaching effectiveness. *College Student Journal, 40,* 195-203.

Isley, P., & Singh, H. (2005). Do higher grades lead to favorable student evaluations? *The Journal of Economic Education, 36,* 29-42.

Johnson, R. L., & Christian, V. K. (1990). Relation of perceived learning and expected grade to rated effectiveness of teaching. *Perceptual and Motor Skills, 70,* 479-482.

Kierstead, D., D'Agostino, P., & Dill, H. (1998). Sex role stereotyping of college professors: Bias in students' ratings of instructors. *Journal of Educational Psychology, 80,* 342-344.

Kinney, D. P., & Smith, S. P. (1992). Age and teaching performance. *The Journal of Higher Education, 63,* 282-302.

Kniepp, L. B., Kelly, K. E., Biscoe, J. D., & Richard, B. (2010). The impact of instructor's personality characteristics on quality of instruction. *College Student Journal, 44,* 901-905.

Koh, H. C., & Tan, T. M. (1997). Empirical investigation of the factors affecting SET results. *International Journal of Educational Management, 11,* 170-178.

Krautmann, A. C., & Sander, W. (1999). Grades and student evaluations of teachers. *Economics of Education Review, 18,* 59-63.

Langbein, L. I. (1994). The validity of student evaluations of teaching. *PS: Political Science and Politics, 27,* 545-553.

Laws, E. L., Apperson, J. M., Buchert, S., & Bregman, N. J. (2010). Student evaluations of instruction: When are enduring first impressions formed? *North American Journal of Psychology, 12,* 81-92.

Liaw, S.-H., & Goh, K.-L. (2003). Evidence and control of biases in student evaluations of teaching. *International Journal of Educational Management, 17,* 37-43.

Marsh, H. W. (1983). Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics. *Journal of Educational Psychology, 75,* 150-166.

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.

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry & J. C. Smart (Eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective,* pp. 319-383, Springer.

Marsh, H. W., Hau, K.-T., Chung, C.-M., & Siu, T. L. P. (1997). Students' evaluations of university teaching: Chinese version of the Students' Evaluations of Educational Quality instrument. *Journal of Educational Psychology, 89,* 568-572.

Marsh, H. W., & Hocevar, D. (1991). Students' evaluations of teaching effectiveness: The stability of mean ratings of the same teachers over a 13-year period. *Teaching and Teacher Education, 7,* 303-314.

Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology, 92,* 202-228.

McCallum, L. W. (1984). A meta-analysis of course evaluation data and its use in the tenure decision. *Research in Higher Education, 21,* 150-158.

McKeachie, W. J. (1996). Do we need norms of student ratings to evaluate faculty? *Instructional Evaluation and Faculty Development, 14,* 14-17.

McPherson, M. A. (2006). Determinants of how students evaluate teachers. *Journal of Economic Education, 37,* 3-20.

McPherson, M. A., & Jewell, R. T. (2007). Leveling the playing field: Should student evaluation scores be adjusted? *Social Science Quarterly, 88,* 868-881.

McPherson, M. A., Jewell, R. T., & Kim, M. (2009). What determines student evaluation scores? A random effects analysis of undergraduate economics classes. *Eastern Economic Journal, 35,* 37-51.

Norvilitis, J. M., & Zhang, J. (2009). The effect of perceived class mean on the evaluation of instruction. *Educational Assessment, Evaluation, and Accountability, 21,* 299-311.

Ory, J. C. (2001). Faculty thoughts and concerns about student ratings. *New Directions for Teaching & Learning, 87,* 3-15.

Patrick, C. L. (2011). Student evaluations of teaching: Effects of the Big Five personality traits, grades and the validity hypothesis. *Assessment & Evaluation in Higher Education, 36,* 239-249.

Perkins, D., Guerin, D., & Schleh, J. (1990). Effects of grading standards information, assigned grade, and grade discrepancies on students' evaluations. *Psychological Reports, 66,* 635-642.

Pounder, J. S. (2007). Is student evaluation of teaching worthwhile?: An analytical framework for answering the question. *Quality Assurance in Education, 15,* 178-191.

Toby, S. (1993). Class size and teaching evaluation: Or, the "general chemistry effect" revisited. *Journal of Chemical Education, 70,* 465-466.

Williams, W. M., & Ceci, S. J. (1997). "How'm I doing?" Problems with student ratings of instructors and courses. *Change: The Magazine of Higher Learning, 29,* 12-23.