



Is student evaluation of teaching worthwhile?

An analytical framework for answering the question

James S. Pounder

Higher Colleges of Technology, Abu Dhabi, United Arab Emirates

Abstract

Purpose – To present a framework to facilitate comprehension of research on the effectiveness of the teaching evaluation process.

Design/methodology/approach – A comprehensive review of the literature that identifies common categories and factors that can be used to construct an analytical framework.

Findings – Identifies student related, course related and teacher related aspects of research on teaching evaluations. Factors commonly addressed within these aspects are also identified.

Research limitations/implications – Use of the framework to analyse the literature on the student evaluation of teaching (SET) process leads to the view that the time is right to explore other methods of assessing classroom dynamics that could supplement the conventional teacher evaluation process.

Practical implications – Educational literature is replete with studies of the SET system, yet due to the preponderance of these studies, it is difficult to take an overview on the effectiveness of this system. On the basis of a comprehensive survey of the literature, this paper identifies and discusses the central factors influencing SET scores. These factors are then presented in a comprehensible table that can be used as a reference point for researchers and practitioners wishing to examine the effectiveness of the SET system.

Originality/value – The paper is one of the few to attempt to make sense of the myriad of studies on teacher evaluation and to develop a framework to facilitate analysis of the effectiveness of the SET system.

Keywords Students, Training evaluation, Classrooms, Leadership

Paper type Research paper

1. Introduction

Student evaluation of teaching (SET) is a widely used instrument in higher education. For example, Seldin (1993) noted an 86 per cent use of the student evaluation of teaching (SET) as a central feature of personnel decisions in US higher education, an increase in usage from 68 per cent in 1984 and 28 per cent in 1973 (Seldin, 1984). In a feature for the *Chronicle of Higher Education*, Wilson (1998, p. A12) stated that:

... only about 30 per cent of colleges and universities asked students to evaluate professors in 1973, but it is hard to find an institution that doesn't today. Such evaluations are now the most important, and sometimes the sole, measure of a teacher's teaching ability.

The extent of reliance on the SET as the predominant measure of university teacher performance is not confined to the USA; it is a worldwide phenomenon (Newton, 1988; Seldin, 1989; Stratton, 1990).

Arguably, the heavy reliance on the SET would be justified if ratings of teacher performance were generally reflected in student achievement. However, there is



considerable disagreement in the literature on the link between SET scores and student achievement. Despite the existence of studies indicating that SET's are reasonably valid multidimensional measures (Marsh and Roche, 1997; McKeachie, 1987) and have a moderate correlation with student learning (d'Apollonia and Abrami, 1997), by and large, most investigations have found little correlation between student achievement and student ratings of their teachers. Cohen's (1983) meta-analysis, for example, found that student achievement accounted for only 14.4 per cent of overall teacher rating variance. Similarly, a meta-analysis by McCallum (1984) found that student achievement explained only 10.1 per cent of overall teacher rating variance. Equally, a 1982 investigation by Dowell and Neal revealed that student achievement accounted for only 3.9 per cent of between-teacher student rating variance (Dowell and Neal, 1982). Finally, a comprehensive study by Damron (1996) found that most of the factors contributing to student ratings of university teachers are probably unrelated to a teacher's ability to promote student learning.

It is findings such as those presented above that have led commentators such as Reckers (1995, p. 33) to state that:

... nearly 75 per cent of academics judge student course evaluations as unreliable and imprecise metrics of performance, yet nearly 100 per cent of schools use them, frequently exclusively.

The remainder of this paper presents a framework for examining research on the factors influencing SET scores that lends support to the view that the typical SET system is seriously flawed.

2. A framework for analysis

The literature is replete with studies of the SET phenomenon (Wilson, 1998) and analysis of the findings indicates a triad comprising student related factors, course related factors, and teacher related factors. This triad is presented in summary form in Table I and is followed by a description of the various factors within the student related, course related and teacher related categories. Arguably, the literature on teacher evaluation generally falls within one or more of these categories and tends to address one or more of the factors subsumed within these categories. Consequently, Table I presents a useful framework for making sense of the myriad of research studies on the SET system.

2.1. Student related factors

Studies tend to revolve around student gender in terms of the extent to which male or female students generally give higher or lower SET scores. Additionally, a few studies have examined the effect of student academic level and maturity on SET scoring. Further, one study has suggested that students use the SET to punish teachers who are perceived to be working them too hard or who have given them low grades. Each of these factors is discussed in more detail below.

Gender effect. More than one study has indicated that student ratings of teachers are influenced by student gender. For example, the study of Walumbwa and Ojode (2000), carried out in a US university, indicated that females, particularly at the undergraduate level, rated their classroom teachers generally higher on classroom leadership dimensions than did their male counterparts. Bachen *et al.* (1999) found a strong

Table I.
Factors influencing SET
scores

<i>Student related factors</i>		
Gender	Generally higher SET scores by male or female students	Bachen <i>et al.</i> (1999); Walumbwa and Ojode (2000); Feldman (1993)
Academic level and maturity	SET score related to academic level of course or student maturity	Aleamoni (1981); Frey <i>et al.</i> (1975); Holtfreter (1991); Langbein (1994); Marsh (1984) Crumbley <i>et al.</i> (2001)
Punishing teachers for low grades		
<i>Course related factors</i>		
Grading	High SET scores for high grades or high grade expectations	Aronson and Linder (1965); Brown (1976); Centra and Creech (1976); Goldman (1993); Greenwald (1997); Johnson and Christian (1990); Perkins <i>et al.</i> (1990) Feldman (1984); Glass <i>et al.</i> (1981); Holtfreter (1991); Koh and Tan (1997); Liaw and Goh (2003); Langbein (1994); Marsh (1987); Meredith (1984); Toby (1993) Cashin (1990); Clark (1993); Cranton and Smith (1986); Aleamoni (1989); DeBerg and Wilson (1990); Stodolsky (1984)
Class size		
Course content	SET scores influenced by academic discipline, degree of course difficulty, required versus elective	
Class timing	SET scores influenced by timing of teaching evaluation when timing of evaluation depends on timing of the course	Cronin and Capie (1986); DeBerg and Wilson (1990); Husbands and Fosh (1993); Koh and Tan (1997)
<i>Teacher related factors</i>		
Gender	SET scores influenced by the gender of the teacher	Bennett (1982); Cooper <i>et al.</i> (1982); Crawford and MacLeod (1990); Downs and Downs (1993); Feldman (1993); Langbein (1994); Rubin (1981); Sears and Hennessey (1996); Kierstead <i>et al.</i> (1988); Winocur <i>et al.</i> (1989) Clayson (1999); Feldman (1983); Holtfreter (1991); Langbein (1994); Smith and Kinney (1992) Bauer (1996); Crumbley (1995); Crumbley <i>et al.</i> (2001); Emery (1995); Handlin (1996); Martin (1998); Powell (1977); Ryan <i>et al.</i> (1980); Sacks (1996); Hocutt (1987-1988); Simpson and Siguaw (2000); Stumpf and Freedman (1979); Winsor (1977); Worthington and Wong (1979); Yunker and Marlin (1984) Abrami <i>et al.</i> (1982); Cardy and Dobbins (1986); Feldman (1986); Jackson <i>et al.</i> (1999); Nafulin <i>et al.</i> (1973); Williams and Ceci (1997)
Age, experience and rank (of teacher)		
Teachers' influencing tactics	Grade inflation, leniency, bringing food to class on the day of the evaluations etc	
Teachers' behavioural traits	The "likeability" factor	

interaction between student gender and professor gender with female students giving especially high ratings to female professors and comparatively lower ratings to male professors on measures reflecting the qualities of being caring-expressive, interactive, professional-challenging, and organized. By contrast, in the same study, the evaluations by male students of male and female professors did not differ significantly on any of these factors. Bachen *et al.*'s (1999) study confirmed similar findings by Feldman (1993).

Student's academic level and maturity. Frey *et al.* (1975) found that more experienced students were clearly more lenient in their ratings than their younger counterparts. Langbein (1994) suggested that higher level students (i.e. those taking higher level courses) are generally more motivated and discriminating in their evaluation of teaching than lower level students. The implication that SET results will tend to be more favourable for higher level subjects has been confirmed by Marsh (1984) and Holtfreter (1991). Further, Aleamoni's (1981) review of prior research cited eight studies that showed no significant relationship between SET results and student level and 18 studies that reported a positive and significant relationship between these two variables. Furthermore, it is interesting to note that Walumbwa and Ojode's (2000) study, referred to earlier, did reveal differences in sensitivity to classroom leadership qualities between the undergraduate and graduate samples.

Students punishing their teachers via SET scores. It is expected that students will use the SET to reflect back to their teachers and the institutions in question, poor teaching performance. However, Crumbley *et al.* (2001), in their examination of students' perception of the evaluation system, discovered that poor SET scores may reflect as much the inadequacy of student effort as they do the quality of the instruction they have received. Thus, Crumbley *et al.* (2001) found that students will punish their teachers via the SET for being asked embarrassing questions (i.e. questions for which the student has no answer), for being graded hard, for being given quizzes and for being given significant homework. Therefore, the SET can be used as a vehicle for students to punish conscientious educators.

2.2. Course related factors

The central area that has received attention is the relationship between grades expected by, or awarded to students and SET scores. Quite simply, there is a sizeable body of work indicating that SET scores are sensitive to grade levels and in particular expected grade levels. Other course related aspects that continue to interest researchers in terms of their effect on SET scores are class size, the nature of the course (i.e. degree of perceived content difficulty, core or elective course etc), and the timing of course delivery (i.e. end of day/week) insofar as this affects the timing of the evaluations. Details of the research findings on these course related aspects are presented below.

Grading. One of the key course related areas that has been investigated in relation to SET scores is the influence of actual grading and students' expectations of grades on SET's. Perkins *et al.* (1990) concluded there was evidence that SET scores were sensitive to the grades professors assigned although Johnson and Christian (1990) noted that expected grades were more highly correlated than assigned grades with student ratings. Nevertheless, both studies confirmed that students with higher than expected grades gave higher SET scores than those with lower than expected grades. While Brown (1976) found that grades accounted for only 9 per cent of variation in

student ratings, he found that grades were substantially more influential than other factors expected to correlate with student ratings. Greenwald (1997) on the other hand, found that grades distort ratings away from the valid measurement of instructional quality by amounts as much as 20 per cent of ratings variance. Centra and Creech (1976) also found a significant correlation between student grade expectations and SET mean rating scores. Therefore, in practice, students are likely to give high ratings in appreciation of high grades (Aronson and Linder, 1965; Goldman, 1993) or the expectation of high grades irrespective of whether these high grades or expectations actually reflect high academic attainment.

Class size. Student ratings of university teachers have been found to vary with class size (Meredith, 1984; Toby, 1993) and, with a few exceptions (e.g. Langbein, 1994; Marsh, 1987), this is one of the most consistent findings in the literature (Koh and Tan, 1997). In general, smaller class sizes tend to result in better SET scores (Feldman, 1984; Holtfreter, 1991; Koh and Tan, 1997; Liaw and Goh, 2003) probably because the opportunity for teacher-student interaction and rapport is greater in smaller sized classes than larger ones (Glass *et al.*, 1981; Toby, 1993). There is, however, a non-linear relationship between class size and SET scores with both relatively small and relatively large classes receiving better ratings (Feldman, 1984; Holtfreter, 1991).

Course content. Stodolsky (1984) has argued that some courses are more difficult to teach than others and thus, course content is likely to influence SET results. Stodolsky's contention is supported by Clark (1993), DeBerg and Wilson (1990) and Cranton and Smith (1986). In contrast, Langbein (1994), despite noting that there is a general perception that teachers delivering "hard" quantitative subjects are likely to receive lower student ratings than those teaching "soft" qualitative subjects, found no evidence of a significant relationship between type of course and overall teaching ratings. However, in a Singaporean setting, Koh and Tan (1997) found that, in a three-year undergraduate business programme, better SET results were associated with first and third year courses than with second year courses. Student academic level and maturity (discussed above) is given as a possible explanation for the third year SET scores and the authors have offered relative ease of learning introductory courses plus student prior familiarity with course content via pre-university studies as likely explanations of the first year phenomenon. They also noted that the nature of the programme under study could have had a significant influence on their results because the programme required students to undertake a particular specialized field in the second year that could prove challenging and that this might account for the relatively lower SET results for courses taken in the second year.

Cashin (1990) examined very large databases of students' ratings and found significant differences in how students rate teaching across various academic disciplines. Hence, arts and humanities courses tend to receive the highest student ratings, biological and social sciences and health and other professions fall into the medium group, English language and literature and history both fall into the medium-low group with business, economics, computer science, mathematics, the physical sciences and engineering falling in the bottom group. Finally, Aleomoni (1989) observed a rating bias against required courses as opposed to elective courses and noted that the more students in a class taking a required course, the lower the relevant SET score, presumably a feature of the interaction of required course and class size (discussed above).

Class timing. Cronin and Capie (1986) found that teaching evaluation results vary from day to day. Thus, to the extent that evaluations are conducted during the classes in question, the timing of classes is a factor affecting SET results. DeBerg and Wilson (1990) and Husbands and Fosh (1993) have suggested that the time and day a course is taught can affect SET results and in a Singaporean university business school context, Koh and Tan (1997) found that SET's conducted in the later part of the week seemed to result in better teaching evaluations. Koh and Tan have speculated that a more relaxed atmosphere exists towards the end of the week that might have a positive effect on SET scores.

2.3. *Teacher related factors*

A central theme in teacher related research is the effect of teacher gender including the influence of gender role expectations on teaching evaluations. Additionally a teacher related dimension that continues to provide a focus for research is what is termed here and in the framework (Table I), teacher influencing tactics. A particular feature of this, for example, is deliberate grade inflation in order to "court" high SET scores. In discussing course related factors earlier in this paper, it was noted that studies have reported a relationship between grade levels and expected grade levels and SET scores. When this relationship is proactively pursued by teachers via a conscious easing up on grades and coursework, there appears to a kind of "mutual back patting" taking place where the teacher gives a high grade to the student (this grade not necessarily reflecting any real student attainment) and, in return, the student rewards the teacher with a high teacher rating. However, teacher influencing tactics need to be distinguished from what is termed in this paper and in the analytical framework, teacher behavioural traits which is another consistent area of research and can be summarized as the effect on SET scores of teacher "likeability". Finally, other teacher related aspects that have been explored by more than one study are the effect on SET scores of age, experience and rank. What follows is a more detailed description of the teacher related factors.

Gender. A great deal has been written about the affect of teachers' gender on SET results often on the premise that female teachers may be discriminated against in what may still be perceived of as a male dominated profession (Koh and Tan, 1997). However, studies of gender effects on SET results do not support a view that female teachers are consistently discriminated against. Thus, Bennett (1982) found that female teachers were consistently rated as friendlier, having a more positive interpersonal style and possessing greater charisma than their male counterparts. Similarly, female teachers have been rated higher than male teachers on the ability to create a classroom environment that invites participation (Crawford and MacLeod, 1990) and on the fostering of a feeling of closeness and warmth for both male and female students (Sears and Hennessey, 1996). Further, a meta-analysis of gender effect on student evaluations conducted by Feldman (1993) indicated that when significant differences were found, they generally favoured the female teacher.

Research also indicates that student ratings are strongly influenced by gender role expectations and, in general, it appears that teacher behaviour perceived by ratees to be inconsistent with traditional gender roles is penalized in student evaluations (Langbein, 1994). For example, females may be expected to be generally more caring and nurturing than men and if a female teacher does not display such qualities in the

view of her students, she may well be penalized in her ratings. Similarly, males may be expected to be more directive and focused on the task than females and likewise may be penalized in student evaluations because students do not perceive them to be operating as expected. Rubin (1981), for instance, found that nurturing qualities were perceived of as more important for female professors than male professors and openness (fairness) more important for male professors. Similarly, Kierstead *et al.* (1988), in asking students to evaluate an imaginary teacher who was male in half the surveys and female in the other half, found that while warmth and interpersonal contact were viewed as important qualities for both male and female versions, the presence of these qualities only influenced students' evaluations of a notional female teacher. Equally, accessibility outside the classroom and a friendly attitude in the class (indicated by a regular smile) positively influenced evaluations of the imaginary female teacher and had no effect on ratings of the male version in the case of accessibility and, in the case of "the ready smile", reduced students' ratings of the male version.

In general, it appears that a number of traits such as warmth, charisma, accessibility, self-assurance and professionalism are valued across faculty gender (Bennett, 1982; Downs and Downs, 1993) but their influence on SET results tends to reflect gender stereotyping. Thus, female teachers perceived of as warm, charismatic and accessible are likely to be more positively evaluated on these traits than their male counterparts (Bennett, 1982; Cooper *et al.*, 1982, Kierstead *et al.*, 1988). Nevertheless, gender stereotyping of female teachers does not always produce positive results for them. Some studies have indicated that stereotyping may alert raters to a perceived shortcoming based on gender that might result in a severe rating if that shortcoming appears to be evident. Therefore, female teachers may be generally perceived to be less professional (professionalism being perceived of as a male quality) than their male colleagues and if the female teacher does not display such a high standard of professionalism that offsets the perception, the female teacher may incur a more negative rating than might otherwise have been the case (Bennett, 1982; Winocur *et al.*, 1989). In summary, the gender-student evaluation relationship is a complex but nonetheless significant factor influencing SETs.

Age, experience, rank. Smith and Kinney (1992) have suggested that the age of a teacher has an effect on SET scores and that older and more experienced teachers tend to receive more positive student evaluations. Furthermore, Holtfreter (1991) found a positive but weak relationship between the rank of a university teacher and student ratings. However, Feldman's (1983) comprehensive review of studies focusing of the influence of teachers' academic rank, instructional experience and age on SETs was not conclusive. Langbein (1994), on the other hand, did find a significant relationship between instructional experience and student ratings although this relationship was non-linear with experience having a positive effect on evaluations up to a point when the effect then became negative. Contrasting with the findings of Smith and Kinney (1992) and Holtfreter (1991), Clayson (1999) found that student evaluations tended to be negatively correlated with the teacher's age and years of experience. In summary, research has produced mixed results and indicates only a potential relationship between teacher age, experience and rank and student ratings.

Teachers' influencing tactics. Earlier, it was noted that despite the widespread use of the SET as the central measure of university teaching performance, academics have

little confidence in its accuracy (Reckers, 1995). Furthermore, SET results often are a major input to personnel decisions relating to academic staff. This situation encourages university teachers to use various tactics to influence student evaluations, many of which, at best, have little educational value and at worst, are actually detrimental to the educational process. As one study suggests:

This SET system causes professors to manipulate students and students in turn to manipulate teachers (Crumbley *et al.*, 2001).

Central to this manipulation are grades. A number of authors have noted that a common method used by teachers to court popularity is grade inflation and “easing up” on course content, assignments and tests (Bauer, 1996; Crumbley, 1995; Handlin, 1996; Ryan *et al.*, 1980; Sacks, 1996). To put it succinctly, university teachers can buy ratings with grades (Hocutt (1987-1988). In a review of faculty tactics aimed at influencing SET outcomes, Simpson and Siguaw (2000) found that the most significant factor reported by faculty was grading leniency and associated activities such as easy or no exams, unchallenging course material and spoon feeding students on examination content. In brief, many university teachers believe that lenient grading produces higher SET scores and they tend to act on this belief (Martin, 1998; Powell, 1977; Stumpf and Freedman, 1979; Winsor, 1977; Worthington and Wong, 1979; Yunker and Marlin, 1984).

Various other manipulative tactics are reported in the literature, many of them fatuous in an educational sense to say the least. For example, Emery (1995) found in a study of 2,673 students at a major US university that teachers who brought food to class received the highest ratings of teaching effectiveness. Simpson and Siguaw (2000) reported that university teachers perceived a major influencing tactic to be the serving of snacks etc. on the day of the evaluations. Other tactics noted by these authors included consistently letting students out of class early, complimenting the class on its ability immediately before administering the evaluation, administering the evaluation when poor students are absent, having a “fun activity” during the class on the day before the evaluation and remaining in the room during the evaluation. Not all the tactics noted by the authors were as irrelevant to the educational process. Some respondents stated that they provided their students with academic extras such as small, in-class, discussion groups and extra study sessions and others stated that they clearly outlined to their students what teaching and learning should be at university level and highlighted expectations in the syllabus. These academic extras were viewed as means of enhancing evaluations via improving students’ academic performance and influencing student expectations. Despite these more positive approaches to influencing SET outcomes, it is evident that much of what is done by academics to influence student evaluations is of little or no educational value.

Teachers’ behavioural traits. This section is distinguished from the previous section in concentrating on the influence of the more subtle university teachers’ behaviour and character traits on SET’s. This is very different from the above focus on the overt, sometimes cynical actions, used by some academics to positively influence SET results. Studies of the effect of personality variables on student evaluations are limited (Simpson and Siguaw, 2000). However, the research that has been done confirms that the behaviour traits of university teachers have a substantial impact on student evaluations. Thus, Feldman (1986) found that the overall relationship of teacher

personality to student ratings is substantial. Williams and Ceci (1997) also found that student ratings are significantly influenced by the personal characteristics of the teacher. Similarly, Cardy and Dobbins (1986) found that students' "liking" of the teacher significantly influenced teaching evaluations. Clayson's (1999) study confirmed that between 50 per cent and 80 per cent of the total variance of student evaluations could be attributed to personality related variables. In a quantitative study, Jackson *et al.* (1999) found that a university teacher's ability to "get on" with students (rapport) overlapped heavily with more squarely educational factors such as teacher enthusiasm for subject, breadth of subject coverage, group interaction and learning value. An extreme interpretation of the type of findings reported by Jackson *et al.* (1999) would support Abrami *et al.* (1982) argument that personable faculty can receive favourable student ratings regardless of how well they know their subject matter (see, for example, Naflulin *et al.*, 1973). In sum, research indicates that university teachers' behavioural traits have a substantial effect on SET results. Studies have also suggested that these behavioural traits may not necessarily be of any educational value.

3. Conclusion and discussion

The above framework highlights the variety of factors influencing the accuracy of student evaluation of teaching and arguably encompasses the major research areas and themes. It is designed to help the researcher and practitioner make sense of the numerous studies that have focused on the SET phenomenon. Perusal of the factors contained in the framework indicates that, although the SET system has its advocates (see, for example, d'Apollonia and Abrami, 1997; Marsh and Roche, 1997; McKeachie, 1987), by and large, most studies have called into question the value of the SET system. It seems that there are so many variables unrelated to the actual execution of teaching influencing SET scores that they tend to obscure accurate assessment of teaching performance. Equally, SET research has generally failed to demonstrate that there is a concrete relationship between teaching performance and student achievement. Accordingly, analysis of the research using the framework presented in this paper suggests the time seems right to explore other methods of evaluating the quality of the classroom experience that could give a more accurate and comprehensive picture of classroom dynamics. For example, a recent study focused on classroom leadership, a notion broader than teaching, and found that effective classroom leadership stimulates extra effort among students (Pounder, 2005). The classroom leadership notion, for example, has considerable potential given the number of studies linking student effort and student achievement (Carbonaro, 2005; Eskew and Faley, 1988; Johnson *et al.*, 2002).

In conclusion, the title of this paper asked the following question: is student evaluation of teaching worthwhile? The framework presented here suggests that in the case of the SET process in its conventional form, its value is questionable as the sole measure of classroom performance since the quality, richness and diversity of what happens in the typical classroom cannot be captured by the SET process alone. However, in the field of education, measures of classroom effectiveness are essential despite the deficiencies of the conventional SET approach. There are therefore strong grounds for arguing that educational organizations can and should experiment with and develop approaches to assessing classroom dynamics that break from the

conventional SET mold. Educational organizations might then be in the position to supplement the conventional SET with other approaches that have the potential to provide a richer picture, and more equitable assessment, of what happens in the classroom.

References

- Abrami, P.C., Leventhal, L. and Perry, R.P. (1982), "Education seduction", *Review of Educational Research*, Vol. 32, pp. 446-64.
- Aleamoni, L.M. (1981), "Student ratings of instruction", in Millman, J. (Ed.), *Handbook of Teacher Evaluation*, Sage Publications, Newbury Park, CA.
- Aleamoni, L.M. (1989), "Typical faculty concerns about evaluation of teaching", in Aleamoni, L.M. (Ed.), *Techniques for Evaluating and Improving Instruction*, Jossey-Bass, San Francisco, CA.
- Aronson, G. and Linder, D.E. (1965), "Gain and loss of esteem as determinants of interpersonal attractiveness", *Journal of Experimental Social Psychology*, Vol. 1, pp. 156-71.
- Bachen, C.M., McLoughlin, M.M. and Garcia, S.S. (1999), "Assessing the role of gender in college students' evaluations of faculty", *Communication Education*, Vol. 48 No. 3, pp. 193-210.
- Bauer, H.H. (1996), "The new generations: students who don't study", paper presented at The Technological Society at Risk Symposium, Orlando, FL.
- Bennett, S.K. (1982), "Student perceptions of and expectations for male and female instructors: evidence relating to the question of gender bias in teaching evaluation", *Journal of Educational Psychology*, Vol. 74, pp. 170-9.
- Brown, D.L. (1976), "Faculty ratings and student grades: a university-wide multiple regression analysis", *Journal of Educational Psychology*, Vol. 68 No. 5, pp. 573-8.
- Carbonaro, W. (2005), "Tracking students' efforts, and academic achievement", *Sociology of Education*, Vol. 78 No. 1, pp. 27-49.
- Cardy, R.L. and Dobbins, G.H. (1986), "Affect and appraisal accuracy: liking as an integral dimension in evaluating performance", *Journal of Applied Psychology*, Vol. 71, pp. 672-8.
- Cashin, W. (1990), "Students do rate different academic fields differently", in Theall, M. and Franklin, J. (Eds), *Student Ratings of Instruction: Issues for Improving Practice*, Jossey-Bass, San Francisco, CA.
- Centra, J.A. and Creech, F.R. (1976), *The Relationship Between Student, Teacher, and Course Characteristics and Student Ratings of Teacher Effectiveness. SIR Report No. 4*, Educational Testing Service, Princeton, NJ, pp. 24-7.
- Clark, D. (1993), "Teacher evaluation: a review of the literature with implications for educators", Seminar in Elementary Education, California State University, Long Beach, CA, Spring.
- Clayson, D.E. (1999), "Students' evaluation of teaching effectiveness: some implications of stability", *Journal of Marketing Education*, Vol. 21, April, pp. 68-75.
- Cohen, P.A. (1983), "Comment on a selective review of the validity of student ratings of teaching", *Journal of Higher Education*, Vol. 54, pp. 448-58.
- Cooper, P., Stewart, L. and Gudykunst, W.B. (1982), "Relationship with instructor and other variables influencing student evaluations of instruction", *Communication Quarterly*, Vol. 30, pp. 308-15.
- Cranton, P. and Smith, R.A. (1986), "A new look at the effect of course characteristics on student ratings of instruction", *American Educational Research Journal*, Spring, pp. 117-28.

-
- Crawford, M. and MacLeod, M. (1990), "Gender in the college classroom: an assessment of the 'chilly climate' for women", *Sex Roles*, Vol. 23, pp. 101-22.
- Cronin, L. and Capie, W. (1986), "The influence of daily variation in teacher performance on the reliability and validity of assessment data", paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.
- Crumbley, D.L. (1995), "The dysfunctional atmosphere of higher education: games professors play", *Accounting Perspectives*, Vol. 1 No. 1, pp. 27-33.
- Crumbley, L., Henry, B.K. and Kratchman, S.H. (2001), "Students' perceptions of the evaluation of college teaching", *Quality Assurance in Education*, Vol. 9 No. 4, pp. 197-207.
- Damron, J.C. (1996), *Instructor Personality and the Politics of the Classroom*, Douglas College, New Westminster.
- d'Apollonia, S. and Abrami, P.C. (1997), "Navigating student ratings of instruction", *American Psychologist*, Vol. 52 No. 11, pp. 1198-208.
- DeBerg, C.L. and Wilson, J.R. (1990), "An empirical investigation of the potential confounding variables in student evaluation of teaching", *Journal of Accounting Education*, Vol. 8 No. 1, pp. 37-62.
- Dowell, D.A. and Neal, J.A. (1982), "A selective view of the validity of student ratings of teaching", *Journal of Higher Education*, No. 53, pp. 51-62.
- Downs, V.C. and Downs, T.M. (1993), DC, "An exploratory and descriptive study identifying communicative behaviors associated with effective college teaching", paper presented at the Annual meeting of the International Communication Association, Washington, DC.
- Emery, C.R. (1995), *Student Evaluations of Faculty Performance*, Clemson University, Clemson, SC.
- Eskew, R.K. and Faley, R.H. (1988), "Some determinants of student performance in the first college-level financial accounting course", *The Accounting Review*, Vol. 63 No. 1, pp. 137-47.
- Feldman, K.A. (1983), "Seniority and experience of college teachers as related to evaluations they receive from students", *Research in Higher Education*, Vol. 18 No. 1, pp. 3-124.
- Feldman, K.A. (1984), "Class size and college students' evaluations of teachers and courses: a closer look", *Research in Higher Education*, Vol. 21 No. 1, pp. 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*, Vol. 24, pp. 139-213.
- 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*, Vol. 34, pp. 151-91.
- Frey, P.W., Leonard, D.W. and Beatty, W.M. (1975), "Student ratings of instruction: validation research", *American Educational Research Journal*, Vol. 12 No. 4, pp. 435-47.
- Glass, G.V., McGaw, B. and Smith, M.L. (1981), *Meta-Analysis in Social Research*, Sage, Beverly Hills, CA.
- Goldman, L. (1993), "On the erosion of education and the eroding foundations of teacher education", *Teacher Education Quarterly*, Vol. 20, pp. 57-64.
- Greenwald, A.G. (1997), "Validity concerns and usefulness of student ratings of instruction", *American Psychologist*, Vol. 52 No. 11, pp. 1182-7.
- Handlin, O. (1996), "A career at Harvard", *American Scholar*, Vol. 65 No. 5, pp. 47-58.
- Hocutt, M.O. (1987-1988), "De-grading student evaluations: what's wrong with student polls of teaching", *Academic Questions*, Winter, pp. 55-64.

-
- Holtfreter, R.E. (1991), "Student rating biases: are faculty fears justified?", *The Woman CPA*, Fall, pp. 59-62.
- Husbands, C.T. and Fosh, P. (1993), "Students' evaluation of teaching in higher education: experiences from four European countries and some implications of the practice", *Assessment and Evaluation in Higher Education*, Vol. 18 No. 2, pp. 95-114.
- Jackson, D.L., Teal, C.R., Raines, S.J. and Nansel, T.R. (1999), "The dimensions of students' perceptions of teaching effectiveness", *Educational and Psychological Measurement*, Vol. 59 No. 4, pp. 580-96.
- Johnson, D.L., Joyce, P. and Sen, S. (2002), "An analysis of student effort and performance in the finance principles course", *Journal of Applied Finance*, Vol. 12 No. 2, pp. 67-72.
- Johnson, R.L. and Christian, V.K. (1990), "Relation of perceived learning and expected grade to rated effectiveness of teaching", *Perceptual and Motor Skills*, Vol. 70, pp. 479-82.
- Kierstead, D., D'Agostino, P. and Dill, H. (1988), "Sex role stereotyping of college professors: bias in student ratings of instructors", *Journal of Educational Psychology*, Vol. 80 No. 3, pp. 342-4.
- Koh, C.H. and Tan, T.M. (1997), "Empirical investigation of the factors affecting SET results", *International Journal of Educational Management*, Vol. 11 No. 4, pp. 170-8.
- Langbein, L.I. (1994), "The validity of student evaluations of teaching", *Political Science and Politics*, September, pp. 545-53.
- Liauw, S.H. and Goh, K.L. (2003), "Evidence and control of biases in student evaluations of teaching", *The International Journal of Educational Management*, Vol. 17 No. 1, pp. 37-43.
- McCallum, L.W. (1984), "A meta-analysis of course evaluation data and its use in the tenure decision", *Research in Higher Education*, Vol. 21, pp. 150-8.
- McKeachie, W. (1987), "Can evaluating instruction improve teaching?", in Aleamoni, L.M. (Ed.), *Techniques for Evaluating and Improving Instruction*, Jossey-Bass, San Francisco, CA.
- Marsh, H.W. (1984), "Students' evaluation of university teaching: dimensionality, reliability, validity, potential biases, and utility", *Journal of Educational Psychology*, October, pp. 707-54.
- Marsh, H.W. (1987), "Students' evaluations of university teaching: research findings, methodological issues, and directions for future research", *Journal of Educational Research*, Vol. 11, pp. 253-388.
- Marsh, H.W. and Roche, L.A. (1997), "Making students' evaluations of teaching effectiveness effective: the critical issues of validity, bias and utility", *American Psychologist*, Vol. 52 No. 11, pp. 1187-97.
- Martin, J.R. (1998), "Evaluating faculty based on student opinions: problems, implications and recommendations from Deming's theory of management perspective", *Issues in Accounting Education*, Vol. 13 No. 4, pp. 1079-94.
- Meredith, G.M. (1984), "Diagnostic and summative appraisal ratings of instruction", *Psychological Reports*, Vol. 46, pp. 21-2.
- Naflulin, D., Ware, J. and Donnelly, F. (1973), "The Dr Fox lecture: a paradigm of educational seduction", *Journal of Medical Education*, Vol. 48, pp. 630-5.
- Newton, J.D. (1988), "Using student evaluation of teaching in administrative control: the validity problem", *Journal of Accounting Education*, Vol. 6 No. 1, pp. 1-14.
- Perkins, D., Gueri, D. and Schleh, J. (1990), "Effects of grading standards information, assigned grade, and grade discrepancies on student evaluations", *Psychological Reports*, Vol. 66, pp. 635-42.

- Pounder, J.S. (2005), "The classroom leadership styles of Hong Kong university teachers: a case study of teachers in a business school", doctoral dissertation, School of Education, Centre for Educational Leadership and Management, University of Leicester.
- Powell, R.W. (1977), "Grades, learning, and student evaluation of instruction", *Research in Higher Education*, Vol. 7, pp. 193-205.
- Reckers, P.M.J. (1995), "Know thy customer", in Baril, C.P. (Ed.), *Change in Accounting Education: A Research Blueprint*, Federation of Schools of Accountancy, St Louis, MO.
- Rubin, R.B. (1981), "Ideal traits and terms of address for male and female college professors", *Journal of Personality and Social Psychology*, Vol. 41, pp. 966-74.
- Ryan, J.I., Anderson, J.A. and Birchler, A.B. (1980), "Evaluations: the faculty responds", *Research in Higher Education*, Vol. 12 No. 4, pp. 317-33.
- Sacks, P. (1996), *Generation X Goes to College*, Open Court, Chicago, IL.
- Sears, S.R. and Hennessey, A.C. (1996), "Students' perceived closeness to professors: the effects of school, professor gender and student gender", *Sex Roles*, Vol. 35, pp. 651-8.
- Seldin, P. (1984), *Changing Practices in Faculty Evaluation*, Jossey-Bass, San Francisco, CA.
- Seldin, P. (1989), "How colleges evaluate professors", *American Association for Higher Education Bulletin*, Vol. 41 No. 7, pp. 3-7.
- Seldin, P. (1993), "The use and abuse of student ratings of professors", *The Chronicle of Higher Education*, Vol. 39 No. 46, p. A40.
- Simpson, P.M. and Siguaw, J.A. (2000), "Student evaluations of teaching: an exploratory study of the faculty response", *Journal of Marketing Education*, Vol. 22 No. 3, pp. 199-213.
- Smith, S.P. and Kinney, D.P. (1992), "Age and teaching performance", *Journal of Higher Education*, Vol. 63 No. 3, pp. 282-302.
- Stodolsky, S. (1984), "Teacher evaluation: the limits of looking", *Educational Researcher*, November, pp. 11-18.
- Stratton, W.O. (1990), "A model for the assessment of student evaluations of teaching, and the professional development of faculty", *The Accounting Educators' Journal*, Summer, pp. 77-101.
- Stumpf, S.A. and Freedman, R.D. (1979), "Expected grade covariation with student ratings of instructors", *Journal of Educational Psychology*, Vol. 71, pp. 273-302.
- Toby, S. (1993), "Class size and teaching evaluation", *Journal of Chemical Education*, Vol. 70 No. 6, pp. 465-6.
- Walumbwa, F.O. and Ojode, L.A. (2000), "Gender stereotype and instructors' leadership behavior: transformational and transactional leadership", paper presented at the Midwest Academy of Management Annual Conference, Chicago, IL, 30 March-1 April.
- Williams, W.M. and Ceci, S.J. (1997), "'How'm I doing?' Problems with student ratings of instructors and courses", *Change*, Vol. 29 No. 5, pp. 12-23.
- Wilson, R. (1998), "New research casts doubt on value of student evaluations of professors", *The Chronicle of Higher Education*, Vol. 44 No. 19, pp. A12-A14.
- Winocur, S., Schoen, L.G. and Sirowatka, A.H. (1989), "Perceptions of male and female academics within a teaching context", *Research in Higher Education*, Vol. 30, pp. 317-29.
- Winsor, J.L. (1977), "A's, B's, but not C's: a comment", *Contemporary Education*, Vol. 48, pp. 82-4.
- Worthington, A.G. and Wong, P.T.P. (1979), "Effects of earned and assigned grades on student evaluations of an instructor", *Journal of Educational Psychology*, Vol. 71, pp. 764-75.

Yunker, J.A. and Marlin, J.W. (1984), "Performance evaluation of college and university faculty: an economic perspective", *Educational Administration Quarterly*, Winter, pp. 9-37.

Further reading

Gellis, Z.D. (2001), "Social work perceptions of transformational and transactional leadership in health care", *Social Work Research*, Vol. 25 No. 1, pp. 17-25.

Michaels, J.W. and Miethe, T.D. (1989), "Academic effort and college grades", *Social Forces*, Vol. 68 No. 1, pp. 309-19.

Naser, K. and Peel, M.J. (1996), "An exploratory study of the impact of intervening variables on student performance in a principles of accounting course", *Accounting Education*, Vol. 7 No. 3, pp. 209-23.

Williams, R.L. and Clark, L. (2002), *Academic Causal Attribution and Course Outcomes for College Students*, ERIC ED 469 337.