Journal of Personnel Evaluation in Education 13:2 153–166, 1999

Student Rating Myths Versus Research Facts from 1924 to 1998

LAWRENCE M. ALEAMONI
Special Education, Rehabilitation, and School Psychology, University of Arizona, Tucson, Arizona 85721

Abstract

Sixteen of the most common myths regarding student ratings of instructors and instruction are looked at from the perspective of the research that has been conducted on them over the past seventy-four years. It is concluded that the myths are, on the whole, myths. However, suggestions are made about how the information regarding the myths can be used to both improve and document instructional effectiveness.

Most of the recent literature on the evaluation of instructional effectiveness has emphasized a need for comprehensive systems. However, a careful scrutiny of actual working systems of instructional evaluation reveals that student ratings of instructor and instruction are still the only component that is regularly obtained and used. Therefore, instructor and instruction evaluations, have become synonymous with student ratings—at least for those being judged. In spite of this widespread use and reliance on student ratings, they remain suspect as a means of evaluating faculty instructional effectiveness. In an attempt to impugn the value of such ratings for faculty self-improvement, promotion, and tenure purposes, faculty and administrators have generated and perpetuated several myths concerning student ratings of instructors and instruction.

To address sixteen of the most common myths regarding student ratings of instructors and instruction, research spanning a seventy-four-year period is cited and summarized below. The research material presented below represents a revision and update of my 1987 article ''Student Rating Myths Versus Research Facts'' published in this Journal.

Myth 1: Students Cannot make Consistent Judgments About the Instructor and Instruction Because of their Immaturity, Lack of Experience, and Capriciousness

Evidence dating back to 1924, according to Guthrie (1954) and more recent literature by Albanese (1991), Hativa (1996), and Palchik (1988) indicate just the opposite. The stability of student ratings from one year to the next resulted in substantial correlations in the range of 0.87 to 0.89. Other literature on the subject, cited by Costin, Greenough, & Menges (1971), and studies by Gillmore (1973) and Hogan (1973) indicated that the correlation between student ratings of the same instructors and courses ranged from 0.70 to 0.87.

Myth 2: Only Colleagues with Excellent Publication Records and Expertise are Qualified to Teach and to Evaluate their Peers' Instruction

A widely held belief (Borgatta, 1970; Deming, 1972) says that good instruction and good research are so closely allied that it is unnecessary to evaluate them independently. Research is divided on this point. Weak positive correlations between research productivity and teaching effectiveness have been found by Maslow & Zimmerman (1956), McDaniel & Feldhusen (1970), McGrath (1962), Riley, Ryan & Lipschitz (1950) and Stallings & Singhal (1968). In contrast, Aleamoni & Yimer (1973) Guthrie (1949, 1954), Hayes (1971), Linsky & Straus (1975), Melland (1996), and Voek (1962) found no significant relationship between instructors' research productivity and students' ratings of their teaching effectiveness. One study Aleamoni & Yimer (1974) also reported no significant relationship between instructors' research productivity and colleagues' ratings of their teaching effectiveness.

Myth 3: Most Student Rating Schemes are Nothing More than a Popularity Contest with the Warm, Friendly, Humorous Instructor Emerging as the Winner Every Time

Studies conducted by Aleamoni & Spencer (1973) which developed and used the Illinois Course Evaluation Questionnaire (CEQ) subscales, indicated that no single subscale (such as Method of Instruction) completely overlapped the other subscales. This result meant that an instructor who received a high rating on the Instructor subscale (made up of items such as ''The instructor seemed to be interested in students as persons'') would not be guaranteed high ratings on the other four subscales (General Course Attitude, Method of Instruction, Course Content, and Interest and Attention). In reviewing both written and objective student comments, Aleamoni (1976) found that students praised instructors for their warm, friendly, humorous manner in the classroom but frankly criticized them if their courses were not well organized or their methods of stimulating students to learn were poor. In fact, Feldman (1989) and Tang (1997) pointed out that students were using preparation and organization, stimulation of interest, motivation, answering questions, and treating students courteously as the basis for their ratings. This evidence, in addition to that presented by Beatty & Zahn (1990), Benz & Blatt (1995), Costin, Greenough & Menges (1971), Dukes & Victoria (1989), Frey (1978), Grush & Costin (1975), Johannessen (1997), Krehbiel et al. (1997), Macdonald (1987), Marlin (1987), Marsh & Bailey (1993), Perry, Abrami & Leventhal (1979), Rodabaugh & Kravitz (1994), Shepherd & Trank (1989), Tollefson, Chen & Kleinsasser (1989), Ware & Williams (1977), and Waters, Kemp & Pucci (1988), indicates that students are discriminating judges of instructional effectiveness.

Myth 4: Students are not able to make Accurate Judgments Until they have been away from the Course and Possibly away from the University for Several Years

It is very difficult to obtain a comparative and representative sample in longitudinal followup studies. The sampling problem is further compounded by the fact that almost all student attitudinal data relating to a course or instructor are gathered anonymously. Most studies in this area, therefore, have relied on surveys of alumni or graduating seniors. Early studies by Drucker & Remmers (1951) showed that alumni who had been out of school five to ten years rated instructors much the same as students currently enrolled. More recent evidence by Aleamoni & Yimer (1974) Marsh (1977), Marsh & Overall (1979), McKeachie, Lin & Mendelson (1978) further substantiated the earlier findings.

Myth 5: Student Rating forms are both Unreliable and Invalid

Well-developed instruments and procedures for their administration can yield high internal consistency reliabilities. Arubayi (1987), Costin, Greenough and Menges (1971), and Marsh (1984) reported such reliabilities to be in the 0.90 range. Aleamoni, 1978a) reported reliabilities ranging from 0.81 to 0.94 for items and from 0.88 to 0.98 for subscales of the CIEQ. It should be noted, however, that wherever student rating forms are not carefully constructed with the aid of professionals, as in the case of most student- and faculty-generated forms Everly & Aleamoni (1972), the reliabilities may be so low as to negate completely the evaluation effect and its results.

Validity is much more difficult to assess than reliability. Most student rating forms have been validated by the judgment of experts that the items and subscales measure important aspects of instruction Costin, Greenough & Menges (1971). These subjectively determined dimensions of instructional setting and process have also been validated using statistical tools, such as factor analysis Aleamoni & Hexner (1980); Burdsal & Bardo (1986); Ellett, Loup & Culross (1997); Marsh (1984). Further evidence of validity comes from studies in which student ratings are correlated with other indicators of teacher competence, such as peer (colleague) ratings, self-ratings, expert judges' ratings, graduating seniors' and alumni ratings, and student learning Abrami, d'Apollonia & Cohen (1990); Baird (1987); Cohen (1989); Dickinson (1990); Drews, Burroughs & Nokovich (1987); Gigliotti & Buchtel (1990); Harrison, Ryan & Moore (1996); Koon & Murray (1995); Nimmer & Stone (1991); O'Connell & Dickinson (1993); Prave & Baril (1993); Prosser & Trigwell (1990); Ryan & Harrison (1995); Shmanske (1988); Stroh (1991); Teven & McCroskey (1997); Vu, Marriott, Skeff, Stratos & Litzelman (1997). In the fourteen studies cited by Aleamoni & Hexner, (1980), student ratings were compared to (1) colleague rating, (2) expert judges' ratings, (3) graduating seniors' and alumni ratings, and (4) student learning measures. All indicated the existence of moderate to high positive correlations, which can be considered as providing additional evidence of validity. This is in contrast to two studies Bendig (1953); Rodin & Rodin (1972) that found a negative relationship between student achievement and instructor rating. The

latter study has been soundly criticized for its methodology by several researchers (Centra, 1973a; Frey, 1973; Gessner, 1973; Menges, 1973).

Myth 6: The Size of the Class affects Student Ratings

Faculty members frequently suggest that instructors of large classes may receive lower ratings because students generally prefer small classes, which permit more student-instructor interaction. Although this belief is supported to some extent by the results of eight studies cited by Aleamoni & Hexner (1980) and two studies cited by Mateo & Fernandez (1996) and Watkins (1990), other investigations do not support it. For example, Aleamoni & Hexner (1980) cited seven other studies that found no relationship between class size and student ratings. Studies cited by Lin (1992) and Van Arsdale & Hammons (1995) as well as one conducted by Shapiro (1990) also indicate no relationship between class size and student ratings. Some investigators have also reported curvilinear relationships between class size and student ratings Gage (1961); Kohlan (1973); Lovell & Haner (1955); Marsh, Overall & Kesler (1979); Pohlmann (1975); Wood, Linsky & Straus (1974).

Myth 7: The Gender of the Student and the Gender of the Instructor Affect Student Ratings

Although conflicting results have been obtained when relating the gender of the student and the gender of the instructor to student evaluations of instruction, a majority of the studies by Aleamoni & Thomas (1980), Amin (1994), Doyle & Whitely (1974), Dukes & Victoria (1989), Feldman (1993), Fernandez & Mateo (1997), Freeman (1994), Goodhartz (1948), Goodwin & Stevens (1993), Hancock, Shannon & Trentham (1992), Isaacson et al. (1964), Ludwig & Meacham (1997), O'Reilly (1987), Petchers & Chow (1988), Wheeless & Potorti (1989), and Winocur, Schoen & Sirowatka (1989) reported no differences between faculty ratings made by male and female students. In addition, Costin, Greenough & Menges (1971) cited seven studies that reported no differences in overall ratings of instructors made by male and female students or in the ratings received by male and female instructors. Conversely, Bendig (1952) found female students to be more critical of male instructors than their male counterparts. Basow & Silberg (1987) and Summers, Anderson, Hines, Gelder & Dean (1996) reported that male and female students both rate female instructors lower than male instructors. Kierstead, D'Agostino & Dill (1988) and Tatro (1995) reported that female instructors are rated higher than males. Walker (1969) found that female students rated female instructors significantly higher than they rated male instructors, whereas Atamian & Ganguli (1993), Goldberg & Callahan (1991), and Luek (1993) reported that male students rate male instructors higher. In addition, Aleamoni & Hexner (1980) cited five studies that reported female students rate instructors higher on some subscales of instructor evaluation forms than do male students.

Myth 8: The Time of Day the Course is Offered Affects Student Ratings

The limited amount of research in this area (Feldman, 1978; Guthrie, 1954; Yongkittikul, Gilmore & Brandenburg, 1974) indicates that the time of day the course is offered does not influence student ratings.

Myth 9: Whether Students take the Course as a Requirement or as an Elective Affects their Ratings

Several investigators have found that students who are required to take a course tend to rate it lower than students who elect to take it (Cohen & Humphreys, 1960; Divoky & Rathermel, 1988; Gillmore & Brandenburg, 1974; Pohlmann, 1975). This finding is supported by Gage (1961), Lovell & Haner (1955), Petchers & Chow (1988), and Scherr & Scherr (1990), who found that instructors of elective courses were rated significantly higher than instructors of required courses. In contrast, Heilman & Armentrout (1936) and Hildebrand, Wilson & Dienst (1971) reported no differences between students' ratings of required courses and elective courses.

Myth 10: Whether Students are Majors or Nonmajors affects their Ratings

The limited amount of research in this area (Aleamoni & Thomas, 1980; Cohen & Humphreys, 1960; Divoky & Rothermel, 1988; Null & Nicholson, 1972; Rayder, 1968) indicates that no significant differences and no significant relationships exist between student ratings and whether students were taking a course as part of a major.

Myth 11: The Level of the Course (Freshman, Sophomore, Junior, Senior, Graduate) Affects Student Ratings

Aleamoni & Hexner (1980) cited eight investigators who reported no significant relationship between student status (freshman, sophomore, etc.) and ratings assigned to instructors. However, they also cited eighteen other investigators who reported that graduate students and upper-division students tended to rate instructors more favorably than did lower division students. More recent studies by (Donaldson, Flannery & Ross-Gordon (1993), Conran (1991), Goldberg & Callahan (1991), and Moritsch & Suter (1988) confirm the differences in ratings by the level of the course on student.

Myth 12: The Rank of the Instructor (Instructor, Assistant Professor, Associate Professor, Professor) affects Student Ratings

Some investigators reported that instructors of higher rank receive higher student ratings (Clark & Keller, 1954; Downie, 1952; Gage, 1961; Guthrie, 1954; Walker, 1969), whereas Schuckman (1990) reported that teaching assistants were rated higher than the ranked faculty. Others reported no significant relationship between instructor rank and student ratings (Aleamoni & Graham, 1974; Aleamoni & Thomas, 1980; Aleamoni & Yimer, 1973; Linsky & Straus, 1975; Petchers & Chow, 1988); Singhal (1968). Conflicting results have also been found when comparing teaching experience to student ratings. Rayder (1968) reported a negative relationship, whereas Heilman & Armentrout (1936) found no significant relationship.

Myth 13: The Grades or Marks Students Receive in the Course are Highly Correlated with their Ratings of the Course and the Instructor

Considerable controversy has centered around the relationship between student ratings and their actual or expected course grades. The general feeling has been that students tend to rate courses and instructors more highly when they expect or receive good grades. Correlational studies have reported widely inconsistent grade-rating relationships. Some twenty-four studies have reported zero relationships (Aleamoni & Hexner, 1980; Baird, 1987; Gigliotti & Buchtel, 1990). Another thirty-seven studies have reported significant positive relationships (Aleamoni & Hexner, 1980; Blunt 1991; Cohen, 1989; Goldberg & Callahan, 1991; Nimmer & Stone, 1991; Rodabaugh & Kravitz, 1994; Sailor, Worthen & Shin, 1997; Scherr & Scherr, 1990; Trick, 1993; Wilson, 1998). One study by Sailor, Worthen & Shin (1997) reported a negative relationship in graduate courses. In most instances, however, these relationships were relatively weak, as indicated by the fact that the median correlation was approximately 0.14, with the mean and standard deviation being 0.18 and 0.16, respectively.

A widely publicized study by Rodin & Rodin (1972) reported a high negative relationship between student performance on examinations and their ratings of graduate teaching assistants. These results have been contested on methodological grounds by Rodin, Frey & Gessner (1975). Subsequent replications of the study using regular faculty rather than teaching assistants and using more sophisticated rating forms have resulted in a positive rather than a negative relationship (Frey, 1973; Gessner, 1973; Sullivan & Skanes, 1974).

Myth 14: There are no Disciplinary Differences in Student Ratings

Studies by Andrew, Gauthier & Jelmberg (1993), Cashin (1990), Goldman (1993), Goodwin & Stevens (1993), and Zahn & Schramm (1992) indicate that disciplinary

differences exist in student ratings. Ratings tend to be higher for the humanities and social science disciplines as compared to the physical science and engineering disciplines.

Myth 15: Student Ratings on Single General Items are Accurate Measures of Instructional Effectiveness

The limited amount of research in this area (Aleamoni & Thomas, 1980; Burdsal & Bardo, 1986; Cashin & Downey, 1992; McBean, 1991; McBean & Lennox, 1987) indicates that there is a low relationship between single general items and specific items and that the single general items had a much higher relationship to descriptive variables (gender, status, required-elective, etc.) than did the specific items. These findings suggest that the use of single general items should be avoided especially for tenure, promotion, or salary considerations.

Myth 16: Student Ratings Cannot Meaningfully be used to Improve Instruction

Studies by Braunstein, Klein & Pachla (1973), Centra (1973b), and Miller (1971) were inconclusive with respect to the effect of feedback at midterm to instructors whose instruction was again evaluated at the end of the term. However, L'Hommedieu, Menges & Brinko (1990), Marsh (1987), Marsh, Fleiner & Thomas (1975), Marsh & Roche (1993), Overall & Marsh (1979), and Sherman (1978) reported more favorable ratings from and improved learning by students by the end of the term. To determine if a combination of a printed report of the results and personal consultations would be superior to providing only a printed report of results, Aleamoni (1978b), Arubayi (1987), Cohen (1991), McKeachie (1979), Schmelkin & Spencer (1997), Schum & Xindra (1996), and Stevens & Aleamoni (1985) found that instructors significantly improved their ratings when personal consultations were provided.

Conclusion

Seventy-three articles were cited in the 1987 article addressing the fifteen most common student rating myths. In the twelve years since that article was published, an additional eighty-one articles have been generated that have addressed the original fifteen student rating myths plus an additional one included in the current article. Amazingly, all of this research continues to support the fact that the sixteen stated student rating myths are (on the whole) myths. In fact, the most recent research on the use of a printed report of the student rating results in combination with personal consultations continue to support the position that such information can be used by the instructor to enrich and improve the course as well as to document instructional effectiveness for administrative purposes.

This review also emphasizes that when student ratings are gathered from forms that have been constructed with the aid of questionnaire design and scaling professionals, the

resulting data and subsequent evaluations can be both reliable and valid. Such data can provide the instructor with first-hand information on the accomplishment of particular educational goals and on the level of satisfaction with and influence of various course elements. If the instructor uses this information constructively, then the students can benefit through an improved teaching and learning environment as well as from having access to information about particular instructors and courses. In addition, administrators (department heads, and deans) also benefit through an improved teaching and learning environment as well as a more accurate representation of student judgments in the personnel decision-making process.

The disadvantages of gathering student ratings primarily result from how they are misinterpreted and misused. The most common misuse is to report raw numerical results and written comments assuming that the user is qualified to interpret such results validly. Without normative (or comparative) information, a faculty member will be tempted to place inappropriate emphasis on selected student responses. If the results are published by the student government association, the biases of the editor(s) might misrepresent the meaning of the ratings to both students and faculty. If administrators use the ratings for punitive purposes only, the faculty will find ways to undermine their use and impugn their credibility.

Acknowledgments

This material represents a revision and update of my 1987 article ''Student Rating Myths Versus Research Facts'' *Journal of Personnel Evaluation in Education, 1*, 111–119.

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 know. *Journal of Educational Psychology, 82*(2), 219–231.

Albanese, M.A. (1991). The validity of lecturer ratings by students and trained observers. *Academic Medicine*, *66*(1), 26–28.

Aleamoni, L.M. (1976). Typical faculty concerns about student evaluation of instruction. *National Association of Colleges and Teachers of Agriculture Journal*, *20*(1), 16–21.

Aleamoni, L.M. (1978a). Development and factorial validation of the Arizona Course/Instructor Evaluation Questionnaire. *Educational and Psychological Measurement*, *38*, 1063–1067.

Aleamoni, L.M. (1978b). The usefulness of student evaluations in improving college teaching. *Instructional Science*, *7*, 95–105.

Aleamoni, L.M., & Yimer, M. (1974). *Graduating senior ratings' relationship to colleague rating, student rating, research productivity and academic rank in rating instructional effectiveness* (Research Report No. 352). Urbana: University of Illinois, Office of instructional Resources, Measurement and Research Division.

Aleamoni, L.M., & Graham, M.H. (1974). The relationship between CEO ratings and instructor's rank, class size, and course level. *Journal of Educational Measurement*, *11*, 189–202.

Aleamoni, L.M., & Hexner, P.Z. (1980). A review of the research on student evaluation and a report on the effect of different sets of instructions on student course and instructor evaluation. *Instructional Science*, *9*, 67–84.

Aleamoni, L.M., & Spencer, R.E. (1973). The Illinois Course Evaluation Questionnaire: A description of its development and a report of some of its results. *Educational and Psychological Measurement*, *33*, 669–684.

Aleamoni, L.M., & Thomas, G.S. (1980). Differential relationships of student, instructor, and course characteristics to general and specific items on a course evaluation questionnaire. *Teaching of Psychology*, *7*(4), 233–235.

Aleamoni, L.M., & Yimer, M. (1973). An investigation of the relationship between colleague rating, student rating, research productivity, and academic rank in rating instructional effectiveness. *Journal of Educational Psychology*, *64*, 274–277.

Amin, M.E. (1994). Gender as a discriminating factor in the evaluation of teaching. *Assessment and Evaluation in Higher Education*, *19*(2), 135–143.

Andrew, M.D., Gauthier, S.A., & Jelmberg, J.R. (1993). Comparing student perceptions of instruction in teacher education and on education courses. *Journal of Personnel Evaluation in Education*, *6*(4), 359–366.

Arubayi, E.A. (1987). Improvement of instruction and teacher effectiveness: Are student ratings reliable and valid? *Higher Education*, *16*(3), 26–278.

Atamian, R., & Ganguli, G. (1993). Teacher popularity and teaching effectiveness: Viewpoint of accounting students. *Journal of Education for Business*, *68*(3), 163–169.

Baird, J.S., Jr. (1987). Perceived learning in relation to student evaluation to university instruction. *Journal of Educational Psychology*, *79*(1), 90–91.

Basow, S.A., & Silverg, N.T. (1987). Student evaluations of college professors: Are female and male professors rated differently? *Journal of Educational Psychology*, *79*(3), 308–314.

Beatty, M.J., & Zahn, C.J. (1990). Are student ratings of communication instructors due to ''easy'' grading practices? An analysis of teacher credibility and student-reported performance levels. *Communication Education*, *39*(4), 275–282.

Bendig, A.W. (1952). A preliminary study of the effect of academic level. sex, and course variables on student rating of psychology instructors. *Journal of Psychology*, *34*, 2–126.

Bendig, A.W. (1953). Relation of level of course achievement of students, instructor and course ratings in introductory psychology. *Educational and Psychological Measurement*, *13*, 437–488.

Benz, C., & Blatt, S.J. (1995). Factors underlying effective college teaching: What students tell us. *Mid-Western Educational Researcher*, *8*(1), 27–31.

Blunt, A. (1991). The effects of anonymity and manipulated grades on student ratings of instructors. *Community College Review*, *18*(4), 48–54.

Borgatta, E.F. (1970). *Student ratings of faculty.* American Association of University Professors, Bulletin, *56*, 6–7.

Braunstein, D.N., Klein, G.A., & Pachla, M. (1973). Feedback, expectancy and shifts in student ratings of college faculty. *Journal of Applied Psychology*, *58*, 254–258.

Burdsal, C.A., & Bardo, J.W. (1986). Measuring student's perceptions of teaching: Dimensions of evaluation. *Educational and Psychological Measurement*, *46*, 63–79.

Cashin, W.E. (1990). Students do rate academic fields differently. *New Directions for Teaching and Learning (Student Ratings of Instruction: Issues for Improving Practice)*, *43*, 113–121.

Cashin, W.E., & Downey, R.G. (1992). Using global student rating items for summative evaluation. *Journal of Educational Psychology*, *84*(4), 563–572.

Centra, J.A. (1973a). Effectiveness of student feedback in modifying college instruction. *Journal of Educational Psychology*, *65*, 395–401.

Centra, J.A. (1973b). The student as godfather? The impact of student ratings on academia. In A.L. Sockloff (ed.), *Proceedings of the First Invitational Conference on Faculty Effectiveness as Evaluated by Students*. Philadelphia: Temple University, Measurement and Research Center.

Clark, K.E., & Keller, R.J. (1954). Student ratings of college teaching. In R.A. Eckert (ed.), *A university looks at its program.* Minneapolis: University of Minnesota Press.

Cohen, J., & Humphreys, L.G. (1960). Memorandum to faculty (unpublished manuscript). University of Illinois, Department of Psychology.

Cohen, P.A. (1989). Do grades influence students' evaluations of clinical courses? *Journal of Dental Education*, *53*(4), 238–240.

Cohen, P.A. (1991). Effectiveness of student rating feedback and consultation for improving instruction in dental schools. *Journal of Dental Education*, *55*(2), 145–150.

Conran, P.B. (1991). High school student evaluation of student teachers: How do they compare with professionals? *Illinois School Research and Development*, *27*(2), 81–92.

Costin, F., Greenough, W.T., & Menges, R.J. (1971). Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research*, *41*, 511–535.

Deming, W.E. (1972). Memorandum on teaching. *The American Statistician*, *26*, 47.

Dickinson, D.J. (1990). The relationship between ratings of teacher performance and student learning. *Contemporary Educational Psychology*, *15*(2), 142–151.

Divoky, J.J., & Rothermel, M.A. (1988). Student perceptions of the relative importance of dimensions of teaching performance across type of class. *Educational Research Quarterly*, *12*(3), 40–45.

Donaldson, J.F., Flannery, D., & Ross-Gordon, J. (1993). A triangulated study comparing adult college students' perceptions of effective teaching with those of traditional students. *Continuing Higher Education Review*, *57*(3), 147–165.

Downie, N.E.W. (1952). Student evaluation of faculty. *Journal of Higher Education*, *23*, 495–496, 503.

Doyle, K.O., & Whitely, S.E. (1974). Student ratings as criteria for effective teaching. *American Educational Research Journal*, *11*, 259–274.

Drews, D.R., Burroughs, W.J., & Nokovich, D.A. (1987). Teacher self-ratings as a validity criterion for student evaluations. *Teaching of Psychology*, *14*(1), 23–25.

Drucker, A.J., & Remers, H.H. (1951). Do alumni and students differ in their attitudes toward instructors? *Journal of Educational Psychology*, *42*, 129–143.

Dukes, R.L., & Victoria, G. (1989). The effects of gender, status, and effective teaching on the evaluation of college instruction. *Teaching Sociology*, *17*(4), 447–457.

Ellet, C.D., Loup, K.S., & Culross, R.R. (1997). Assessing enhancement of learning, personal learning environment, and student efficacy: Alterative to traditional faculty evaluation in higher education. *Journal of Personnel Evaluation in Education*, *11*(2), 167–192.

Everly, J.C., & Aleamoni, L.M. (1972). The rise and fall of the advisor: Students attempt to evaluate their instructors. *Journal of the National Association of Colleges and Teachers of Agriculture*, *16*(2), 43–45.

Feldman, K.A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education*, *9*, 199–242.

Feldman, K.A. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education*, *30*(6), 583–645.

Feldman, K.A. (1993). College students' views of male and female college teachers: Part II, Evidence from students' evaluations of their classroom teachers. *Research in Higher Education*, *34*(2), 151–211.

Fernandez, J., & Mateo, M.A. (1997). Student and faculty gender in rating of university teaching quality. *Sex Roles: A Journal of Research*, *37*(11–12), 997–1003.

Freeman, H.R. (1994). Student evaluations of college instructors: Effects of type of course taught, instructor gender and gender role, and student gender. *Journal of Educational Psychology*, *86*(4), 627–630.

Frey, P.W. (1973). Student ratings of teaching: Validity of several rating factors. *Science*, *182*, 83–85.

Frey, P.W. (1978). A two-dimensional analysis of student ratings of instruction. *Research in Higher Education*, *9*, 69–91.

Gage, N.L. (1961). The appraisal of college teaching. *Journal of Higher Education*, *32*, 17–22.

Gessner, P.K. (1973). Evaluation of instruction. *Science*, *180*, 566–569.

Gigliotti, R.J., & Buchtel, F.S. (1990). Attributional bias and course evaluations. *Journal of Educational Psychology*, *82*(2), 341–351.

Gillmore, G.M. (1973). *Estimates of reliability coefficients for items and subscales of the Illinois Course Evaluation Questionnaire* (Research Report No. 341). Urbana: University of Illinois, Office of Instructional Resources, Measurement, and Research Division.

Gillmore, G.M., & Brandenburg, D.C. (1974). *Would the proportion of students taking a class as a requirement affect the student rating of the course?* (Research Report No. 347). Urbana: University of Illinois, Office of Instructional Resources, Measurement, and Research Division.

Goldberg, G, & Callahan, J. (1991). Objectivity of student evaluations of instructors. *Journal of Education for Business*, *66*(6), 377–378.

Goldman, L. (1993). On the erosion of education and the eroding foundation of teacher education (or why we should not take student evaluation of faculty seriously). *Teacher Education Quarterly*, *20*(2), 57–64.

Goodhartz, A.S. (1948). Student attitudes and opinions relating to teaching at Brooklyn College. *School and Society*, *68*, 345–349.

Goodwin, L.D., & Stevens, E.A. (1993). The influence of gender on university faculty members' perceptions of ''good'' teaching. *Journal of Higher Education*, *64*(2), 166–185.

Grush, J.E., & Costin, F. (1975). The student as consumer of the teaching process. *American Educational Research Journal*, *12*, 55–66.

Guthrie, E.R. (1949). The evaluation of teaching. *Educational Record*, *30*, 109–115.

Guthrie, E.R. (1954). *The evaluation of teaching: A progress report*. Seattle: University of Washington.

Hancock, G.R., Shannon, D.M., & Trentham, L.L. (1992). Student and teacher gender ratings of university faculty: Results from five colleges of study. *Journal of Personnel Evaluation in Education*, *6*(3), 235–248.

Harrison, P.D., Ryan, J.M., & Moore, P.S. (1996). College students' self-insight and common implicit theories in ratings of teaching effectiveness. *Journal of Educational Psychology*, *88*(4), 775–782.

Hativa, N. (1996). University instructors' rating profiles: Stability over time, and disciplinary differences. *Research in Higher Education*, *37*(3), 341–365.

Hayes, J.R. (1971). Research, teaching and faculty fate. *Science*, *172*, 227–230.

Heilman, J.D., & Armentrout, W.D. (1936). The rating of college teachers on ten traits by their students. *Journal of Educational Psychology*, *27*, 197–216.

Hildebrand, M., Wilson, R.C., & Dienst, E.R. (1971). *Evaluating university teaching*. Berkeley: University of California, Center for Research and Development in Higher Education.

Hogan, T.P. (1973). Similarity of student ratings across instructors, courses and time. *Research in Higher Education*, *1*, 149–154.

Isaacson, R.L., McKeachie, W.J., Milholland, J.E., Lin, Y.G., Hotelier, M., Baerwaldt, J.W. & Zinn, K.L. (1964). Dimensions of student evaluations of teaching. *Journal of Educational Psychology*, *55*, 344–351.

Johannessen, T.A. (1997). What is important to students? Exploring dimensions in their evaluations of teachers. *Scandinavian Journal of Educational Research*, *41*(2), 165–177.

Kierstead, D., D'Agostino, P., & Dill, H. (1988). Sex role stereotyping of college professors: Bias in students' ratings of instructors. *Journal of Educational Psychology*, *80*(3), 342–344.

Kohlan, R.G. (1973). A comparison of faculty evaluations early and late in the course. *Journal of Higher Education*, *44*, 587–595.

Koon, J., & Murray, H.G. Using multiple outcomes to validate student ratings of overall teacher effectiveness. *Journal of Higher Education*, *66*(1), 61–81.

Krehbiel, T.C., & associates (1997). Using student disconfirmation as a measure of classroom effectiveness. *Journal of Education for Business*, *72*(4), 224–229.

L'Hommedieu, R., Menges, R.J., & Brinko, K.T. (1990). Methodological explanations for modest effects of feedback from student ratings. *Journal of Educational Psychology*, *82*(2), 232–241.

Lin, W.Y. (1992). Is class size a bias to student ratings of university faculty? A review. *Chinese University of Education Journal*, *20*(1), 49–53.

Linsky, A.S., & Straus, M.A. (1975). Student evaluations, research productivity and eminence of college faculty. *Journal of Higher Education*, *46*, 89–102.

Lovell, G.D., & Haner, C.F. (1955). Forced-choice applied to college faculty rating. *Educational and Psychological Measurement*, *15*, 291–304.

Ludwig, J.M., & Meacham, J.A. (1997). Teaching controversial courses: Student evaluations of instructor and content. *Educational Research Quarterly*, *21*(1), 27–38.

Luek, T.L. (1993). The interaction effects of gender on teaching evaluations. *Journalism Educator*, *48*(3), 235–248.

Macdonald, A. (1987). Student views on excellent courses. *Agricultural Education Magazine*, *60*(3), 19–22.

Marlin, J.W., Jr. (1987). Student perceptions of end-of-course evaluation. *Journal of Higher Education*, *58*(6), 704–716.

Marsh, H.W. (1977). The validity of students' evaluations: Classroom evaluations of instructors independently nominated as best and worst teachers by graduating seniors. *American Educational Research Journal*, *14*, 441–447.

Marsh, H.W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, *76*, 707–754.

Marsh, H.W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, *11*(3), 253–388.

Marsh, H.W., & Bailey, M. (1993). Multidimensional students' evaluation of teaching effectiveness. *Journal of Higher Education*, *64*(1), 1–18.

Marsh, H.W., Fleiner, H., & Thomas, C.S. (1975). Validity and usefulness of student evaluations of instructional quality. *Journal of Educational Psychology*, *67*, 833–839.

Marsh, H.W., Overall, J.U., & Kesler, S.P. (1979). Class size, student evaluations, and instructional effectiveness. *American Educational Research Journal*, *16*, 57–69.

Marsh, H.W., & Overall, J.U. (1979). Long-term stability of students' evaluations: A note of Feldman's consistency and variability among college students in rating their teachers and courses. *Research in Higher Education*, *10*, 139–147.

Marsh, H.W., & Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American Educational Research Journal*, *30*(1), 217–251.

Maslow, A.H., & Zimmerman, W. (1956). College teaching ability, scholarly activity and personality. *Journal of Educational Psychology*, *47*, 185–189.

Mateo, M.A. & Fernandez, J. (1996). Incidence of class size on the evaluation of university teaching quality. *Educational and Psychological Measurement*, *56*(5), 771–778.

McBean, E.A. (1991). Analyses of teaching and course questionnaires: A case study. *Engineering Education*, *81*(4), 439–441.

McBean, E.A., & Lennox, W.C. (1987). Measurement of quality of teaching and course by a single question versus a weighted set. *European Journal of Engineering Education*, *12*(4), 329–335.

McDaniel, E.D., & Feldhusen, J.F. (1970). Relationships between faculty ratings and indexes of service and scholarship. *Proceedings of the Seventy-Eighth Annual Convention of the American Psychological Association*, *5*, 619–620.

McGrath, E.J. (1962). Characteristics of outstanding college teachers. *Journal of Higher Education*, *33*, 148.

McKeachie, W.J. (1979). Student ratings of faculty: A reprise. *Academe*, *65*, 384–397.

McKeachie, W.J., Lin, Y.G., & Mendelson, C.N. (1978). A small study assessing teacher effectiveness: Does learning last? *Contemporary, Educational Psychology*, *3*, 352–357.

Melland, H.I. (1996). Great researcher . . . good teacher? *Journal of Professional Nursing*, *12*(1), 31–38.

Menges, R.J. (1973). The new reporters: Students rate instruction. In C.R. Pace (ed.), *New directions in higher education: evaluating learning and teaching*. San Francisco: Jossey-Bass.

Miller, M.T. (1971). Instructor attitudes toward, and their use of, student ratings of teachers. *Journal of Educational Psychology*, *62*, 235–239.

Moritsch, B.G., & Suter, W.N. (1988). Correlates of halo error in teacher evaluation. *Educational Research Quarterly*, *12*(3), 29–34.

Nimmer, J.G., & Stone, E.F. (1991). Effects of grading practices and time of rating on student ratings of faculty performance and student learning. *Research in Higher Education*, *32*(2), 195–215.

Null, E.J., & Nicholson, E.W. (1972). Personal variables of students and their perception of university instructors. *College Student Journal*, *6*, 6–9.

O'Connell, D.Q., & Dickenson, D.J. (1993). Student ratings of instruction as a function of testing conditions and perceptions of amount learned. *Journal of Research and Development in Education*, *27*(1), 18–23.

O'Reilly, M.T. (1987). Relationship of physical attractiveness to students ratings of teaching effectiveness. *Journal of Dental Education*, *51*(10), 600–602.

Overall, J.U., & Marsh, H.W. (1979). Midterm feedback from students: Its relationship to instructional improvement and students' cognitive and affective outcomes. *Journal of Educational Psychology*, *71*, 856–865.

Palchik, N.S. (1988). Student assessment of teaching effectiveness in a multi-instructor course for multidisciplinary health professional students. *Evaluation and the Health Professions*, *11*(1), 55–73.

Perry, R.P., Abrami, P.C., & Leventhal, L. (1979). Educational seduction: The effect of instructor expressiveness and lecture content on student ratings and achievement. *Journal of Educational Psychology*, *71*, 107–116.

Petchers, M.K., & Chow, J.C. (1988). Sources of variation in students' evaluations of instruction in a graduate social work program. *Journal of Social Work Education*, *24*(1), 35–42.

Pohlmann, J.T. (1975). A multivariate analysis of selected class characteristics and student ratings of instruction. *Multivariate Behavioral Research*, *10*(1), 81–91.

Prave, R.S., & Baril, G.L. (1993). Instructor ratings: Controlling for bias from initial student interest. *Journal of Education for Business*, *68*(6), 362–366.

Prosser, M., & Trigwell, K. (1990). Student evaluations of teaching and courses: Student study strategies as a criterion of validity. *Higher Education*, *20*(2), 135–142.

Rayder, N.F. (1968). College student ratings of instructors. *Journal of Experimental Education*, *37*, 76–81.

Riley, J.W., Ryan, B.F., & Lipschitz, M. (1950). *The student looks at his teacher*. New Brunswick, NJ: Rutgers University Press.

Rodabaugh, R.C., & Kravitz, D.A. (1994). Effects of procedural fairness on student judgements of professors. *Journal on Excellence in College Teaching*, *5*(2), 67–83.

Rodin, M., Frey, P.W., & Gessner, P.K. (1975). Student evaluation. *Science*, *187*, 555–559.

Rodin, M., & Rodin, B. (1972). Student evaluations of teachers. *Science*, *177*, 1164–1166.

Ryan, J.M., & Harrison, P.D. (1995). The relationship between individual instructional characteristics and the overall assessment of teaching effectiveness across different instructional contexts. *Research and Development in Education*, *36*(5), 577–594.

Sailor, P., Worthen, B.R., & Shin, E.H. (1997). Class level as a possible mediator of the relationship between grades and student ratings of teaching. *Assessment and Evaluation in Higher Education*, *22*(3), 261–269.

Scherr, F.C., & Scherr, S.S. (1990). Bias in student evaluation of teacher effectiveness. *Journal of Education for Business*, *65*(8), 356–358.

Schmelkin, L.P., & Spencer, K.J. (1997). Faculty perspectives on course and teacher evaluations. *Research in Higher Education*, *38*(5), 75–92.

Schuckman, H. (1990). Students' perception of faculty and graduate students as classroom teachers. *Teaching of Psychology*, *17*(3), 162–165.

Schum, T.R., & Vindra, K.J. (1996). Relationship between systematic feedback to faculty and ratings of clinical teaching. *Academic Medicine*, *71*(10), 1100–1102.

Shapiro, E.G. (1990). Effect of instructor and class characteristics on students' class evaluations. *Research in Higher Education*, *3*(2), 135–148.

Shepherd, G.J., & Trank, D.M. (1989). Individual differences in consistency of evaluation: Student perceptions of teacher effectiveness. *Journal of Research and Development in Education*, *22*(3), 45–52.

Sherman, T.M. (1978). The effects of student formative evaluation of instruction on teacher behavior. *Journal of Educational Technology Systems*, *6*, 209–217.

Shmanske, S. (1988). On the measurement of teacher effectiveness. *Journal of Economic Education*, *19*(4), 307–314.

Singhal, S. (1968). *Illinois Course Evaluation Questionnaire items by rank of instructor, sex of the instructor and sex of the student* (Research Report No. 282). Urbana: University of Illinois, Office of Instructional Resources, Measurement, and Research Division.

Stallings, W.M., & Singhal, S. (1968). *Some observations on the relationships between productivity and student evaluations of course and teaching* (Research Report No. 274). Urbana: University of Illinois, Office of Instructional Resources, Measurement, and Research Division.

Stevens, J.J., & Aleamoni, L.M. (1985). The use of evaluative feedback for instructional improvement: A longitudinal perspective. *Instructional Science*, *13*, 285–304.

Sullivan, A.M., & Skanes, G.R. (1974). Validity of student evaluation of teaching and the characteristics of successful instructors. *Journal of Educational Psychology*, *66*, 584–590.

Summers, M.A., Anderson, J.L., Hines, A.R., Gelder, B.C., & Dean, R.S. (1996). The camera adds more than pounds: Gender differences in course satisfaction for campus and distance learning students. *Journal of Research and Development in Education*, *29*(4), 212–219.

Tang, T.L.-P. (1997). Teaching evaluation at a public institution of higher education: Factors related to the overall teaching effectiveness. *Public Personnel Management*, *26*(3), 379–389.

Tatro, C.N. (1995). Gender effects on student evaluations of faculty. *Journal of Research and Development in Education*, *28*(3), 169–173.

Teven, J.J., & McCroskey, J.C. (1997). The relationship of perceived teacher caring with student learning and teacher evaluation. *Communications Education*, *46*(1), 1–9.

Tollefson, N., Chen, J.S., & Kleinsasser, A. (1989). The relationship of students' attitudes about effective teaching to students' ratings of effective teaching. *Educational and Psychological Measurement*, *49*(3), 529–536.

Trick, L.R. (1993). Do grades affect faculty teaching evaluations? *Journal of Optometric Education*, *18*(3), 88–92.

Voeks, V.W. (1962). Publications and teaching effectiveness. *Journal of Higher Education*, *33*, 212.

Vu, T.R., Marriott, D.J., Skeff, K.M., Stratos, G.A., & Litzelman, D.K. (1997). Prioritizing areas for faculty development of clinical teachers by using student evaluations for evidence-based decisions. *Academic Medicine*, *72*(10), 57–59.

Walker, B.D. (1969). An investigation of selected variables relative to the manner in which a population of junior college students evaluate their teachers. *Dissertation Abstracts*, *29*(9-B), 3474.

Ware, J.E., & Williams, R.G. (1977). Discriminate analysis of student ratings as a means of identifying lecturers who differ in enthusiasm or information giving. *Educational and Psychological Measurement*, *37*, 627–639.

Waters, M., Kemp, E., & Pucci, A. (1988). High and low faculty evaluations: Descriptions by students. *Teaching of Psychology*, *15*(4), 203–204.

Watkins, D. (1990). Student ratings of tertiary courses for ''alternative calendar'' purposes. *Assessment and Evaluation in Higher Education*, *15*(1), 12–21.

Wheeless, V.E., & Potorti, P.F. (1989). Student assessment of teacher masculinity and femininity: A test of the sex role congruency hypothesis on student attitudes toward learning. *Journal of Educational Psychology*, *81*(2), 259–262.

Wilson, R. (1998). New research casts doubt on value of comparing adult college students perceptions of effective teaching with those of traditional students. *Chronicle of Higher Education*, *44*(19), A12–A14.

Winocur, S., Schoen, L.G., & Sirowatka, A.H. (1989). Perceptions of male and female academics within a teaching context. *Research in Higher Education*, *30*(3), 317–329.

Wood, K., Linsky, A.S., & Straus, M.A. (1974). Class size and student evaluations of faculty. *Journal of Higher Education*, *45*, 524–534.

Yongkittikul, C., Gillmore, G.M., & Brandenburg, D.C. (1974). *Does the time of course meeting affect course ratings by students?* (Research Report No. 346). Urbana: University of Illinois, Office of Instructional Resources, Measurement, and Research Division.

Zahn, D.K., & Schramm, R.M. (1992). Student perception of teacher effectiveness based on teacher employment and course skill level. *Business Education Forum*, *46*(3), 16–18.