Borrower: EMU

Lending String: *KUK,VA@,FWA,MNU,TTU

ILL Number: 58381113

OCLC #: 16164930
Journal Title: Journal of personnel evaluation in education.
Volume: 1 Issue: 1
Month/Year: 1987 Pages: 111-119

Article Author:

Article Title: ALEAMONI,; STUDENT RATING MYTHS VERSUS RESEARCH FACTS

Imprint: Boston/U.S.A. ; Kluwer, 1987-

Call #: LB2838 .J68 - -

Charge
Maxcost: $35IFM
Billing Category: Exempt

Shipping Address:
Emory University
Interlibrary Loan
Robert W. Woodruff Library
540 Asbury Circle, NE
Atlanta, GA 30322-2870
Fax: 404-727-0052
Ariel: 170.140.209.109
Patron: WEEKS, ERIC

Notes:
Student Rating Myths Versus Research Facts

LAWRENCE M. ALEAMONI
Instructional Research and Development, University of Arizona, Tucson, Arizona 85721

Most of the recent literature on the evaluation of instructional effectiveness has emphasized the need to develop comprehensive systems. However, a careful scrutiny of actual working systems of instructional evaluation reveals that student ratings of instructor and instruction is still the only component that is regularly obtained and used. Therefore, instructor/instructional evaluation has become synonymous with student rating/evaluation for those being judged. In an attempt to impugn the value of such ratings for faculty self improvement and/or promotion and tenure purposes, faculty and administrators have generated and perpetuated several myths concerning student ratings of instructors and instruction.

In order to address 15 of the most common myths regarding student ratings of instructors and instruction, research spanning a 62-year period will be cited and summarized below.

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), indicates 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. More recent literature on the subject, cited by Costin, Greenough, and 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.

There is a widely held belief (Borgatta, 1970; Deming, 1972) 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 and Zimmerman (1956), McDaniel and Feldhusen (1970), McGrath (1962), Riley, Ryan, and Lipschitz (1950), and Stallings and Singhal (1968). In contrast, Aleamoni and Yimer (1973), Guthrie (1949, 1954), Hayes (1971). Linsky

and Straus (1975), and Voeks (1962) found no significant relationship between instructors' research productivity and students' ratings of their teaching effectiveness. One study (Aleamoni & Yimer, 1973) 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 and Spencer (1973), while developing and using the Illinois Course Evaluation Questionnaire (CEQ) subscales, indicated that no single subscale (e.g., 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 frankly praised instructors for their warm, friendly, humorous manner in the classroom, but if their courses were not well organized or their methods of stimulating students to learn were poor, the students equally frankly criticized them in those areas. This evidence, in addition to that presented by Costin and associates (1971), Frey (1978), Grush and Costin (1975), Perry, Abrami, and Leventhal (1979), and Ware and Williams (1977), 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 and/or graduating seniors. Early studies by Drucker and Remmers (1951) showed that alumni who had been out of school 5 to 10 years rated instructors much the same as students currently enrolled. More recent evidence by Aleamoni and Yimer (1974), Marsh (1977), Marsh and Overall (1979), McKeachie, Lin, and 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. Costin and associates (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 CEIQ. 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 and 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 with objective criteria only. Despite this, validity has been validated with objective criteria (1980; Builder, 1980; Costin, 1971) from student ratings of the behavior of a teacher or by correlating students' ratings with those given by Aleamoni, Yimer, and colleagues (1973). In these studies students' ratings, and not those of instructors, were related to high psychometric scores. Further evidence of validity is provided by Rodin (1975), who found that a measure of instructor effectiveness, validated by several years of experience, was related to student ratings.

**Myth 6: Faculty members give students lower ratings than do other raters.**

Conflicting conclusions have been drawn as to student expectations. Whitely (1970), Hofeller, Baltimore and associates (1971), and ratings made by faculty and instructors in the same courses (1971) cited as evidence that students rate instructors more favorably than do faculty and female instructors are rated more critically than male instructors. Walker (1965) found that student ratings are higher than those of faculty. There are cited five studies, showing the first subscales of instruments.

**Myth 7: The第一名 student is the best student.**

The limited evidence indicates that Yongkittikul, student raters do not rate course is offered with the best raters.
forms have been validated by the judgment of experts that the items and subscales measure important aspects of instruction (Costin et al., 1971). These subjectively determined dimensions of instructional setting and process have also been validated using statistical tools, such as factor analysis (Aleamoni & Hexner, 1980; Burdahl & Bardo, 1986; 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, expert judges' ratings, graduating seniors' and alumni ratings, and student learning. The 14 studies cited by Aleamoni and Hexner (1980) in which 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, 1973b; 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 and Hexner (1980), other investigations do not support it. For example, Aleamoni and Hexner (1980) cited seven other studies that found 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, & Strauss, 1974).

Myth 7: Gender of the student and the instructor affect student ratings.

Conflicting results have been obtained when relating the gender of the student to student evaluations of instruction. Aleamoni and Thomas (1980), Doyle and Whitely (1974), Goodhart (1948), and Isaacs, McKeachie, Milholland, Lin, Hefler, Baerwaldt, and Zinn (1964) reported no differences between faculty ratings made by male and female students. In addition, Costin and associates (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; more recently Walker (1969) found that female students rated female instructors significantly higher than they rated male instructors. In addition, Aleamoni and 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
affect 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; Gillmore & Brandenburg, 1974; Pohlmann, 1975). This finding is supported by Gage (1961) and Lovell and Haner (1955), who found that instructors of elective courses were rated significantly higher than instructors of required courses. In contrast, Heilman and Armentrout (1936) and Hildebrand, Wilson, and Dienst (1971) reported no differences between students’ ratings of required courses and elective courses.

Myth 10: Whether students are majors or nonmajors affect their ratings.

The limited amount of research in this area (Aleamoni & Thomas, 1980; Cohen & Humphreys, 1960; Null & Nicholson, 1972; Rayder, 1968) indicates that there are no significant differences and no significant relationships between student ratings and whether they were majors or minors.

Myth 11: The level of the course (freshman, sophomore, junior, senior, graduate) affects student ratings.

Aleamoni and 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 18 other investigators who reported that graduate students and/or upper division students tended to rate instructors more favorably than did lower division students.

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); however, others reported no significant relationship between instructor rank and student ratings (Aleamoni & Graham, 1974; Aleamoni & Thomas, 1980; Aleamoni & Yimer, 1973; Linsky & Straus, 1975; Singhal, 1968). Conflicting results have also been found when comparing teaching experience to student ratings. Rayder (1968) reported a negative relationship, whereas Heilman and 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 being 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 22 studies have reported zero relationships (Aleamoni & Hexner 1980). Another 28 studies have reported significant positive relationships (Aleamoni & Hexner, 1980). 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 and 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, and 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: Student ratings on single general items are accurate measures of instructional effectiveness.

The limited amount of research in this area (Aleamoni & Thomas, 1980; Bardsal & Bardo, 1986) 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 15: Student ratings cannot meaningfully be used to improve instruction.

Studies by Braunstein, Klein, and Pachla (1973), Centra (1973a), 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, Marsh, Fleiner, and Thomas (1975), Overall and Marsh (1979), and Sherman (1978) reported more favorable ratings from and improved learning by students by the end of the term. In order 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), McKeachie (1979), and Stevens and Aleamoni (1985) found that instructors significantly improved their ratings when personal consultations were provided.

Conclusion

All this research points out that the previously stated student rating myths are (on the whole) myths. On the other hand, gathering student ratings 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. Such information can be used by the instructor to enrich and improve the course as well as to document instructional effectiveness for administrative purposes.

Students can benefit through an improved teaching and learning situation as well as from having access to information about particular instructors and courses. Administrators (deans and department heads) also benefit through an improved teaching and learning situation as well as a more accurate representation of student judgments.

The disadvantages of gathering student ratings primarily result from how they are misinterpreted and misused. Without normative (or comparative) information, a faculty member might place inappropriate emphasis on selected student
responses. If the results are published, 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 be unfairly represented.

References


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

Costin, F., Greenough, W.T., & Menges, R.J. (1971). Student ratings of college teaching: Reliability,
Duffy, J.C., & Akefanni, 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, 10(2), 43-45.
Goodhardt, A.S. (1948). Student attitudes and opinions relating to teaching at Brooklyn College.
School and Society, 68, 345-349.


