

Developments in Business Simulation & Experiential Exercises, Volume 12, 1985

PROBLEMS ASSOCIATED WITH THE ADMINISTRATIVE USES OF STUDENT RATINGS

Allan S. Baillie, California Polytechnic State University
Robert H. McIntire, California Polytechnic State University

ABSTRACT

It might be postulated that college administrators formulate policy with due regard given to educational research. However, based on a review of the literature covering the administrative uses of student ratings, it appears the opposite may be true. In many instances the administrators are unfamiliar with the pertinent research literature. For example, how many administrators realize that research findings in areas such as: class attendance, class sizes, and the characteristics of good teaching do not support the conventional wisdom dictating policy in these areas? Stephen [1].

INTRODUCTION

There have been hundreds of studies investigating the appropriate use of student ratings, the Biddle annotated bibliography for years 1974 through 1978 lists over 200 studies. Biddle [2]. The results of this massive research effort offer findings to support the use of student ratings to evaluate teaching Beatty [3], Becholt [4], McKeachie [5], Marsh and Overall [6], Perry [7], Sullivan and Skanes [8], and many challenges to their use Benington [9], Bers [10], Bligh [11], Chisholm [12], Cohen [13], Moody [14], O'Tuel [15], and Rodin and Rodin [16]. The researchers in this area seem to agree that student ratings can be useful but that for personnel actions other sources of information must be used in conjunction with the student ratings. Boyd and Schietinger [17] report 88% of the institutions surveyed used student ratings for decision making, yet administrative policy does not reflect the concerns voiced by most researchers because in many institutions (and state college systems) student ratings have become the primary basis for evaluating faculty under consideration for promotion, retention and tenure.

The Dr. Fox Effect

Educational seduction or the "Dr. Fox Effect" is a finding that has been especially troublesome to those advocating the appropriateness of student ratings for faculty personnel actions. In essence the effect assumes that an entertaining, charismatic teacher will receive high student ratings even while offering minimum lecture content. The idea for this line of research was based on the work of Edwin Goffman, who suggested that expressive behavior may influence an audience as much or more than substance when there is little basis for the audience to doubt the credentials of the speaker. Goffman [18].

The original attempt to demonstrate the "Dr. Fox Effect" involved an actor who was an accomplished double talk artist. He was armed with a lecture written in an entertaining style but reported a fabricated research project ("Mathematical Game Theory Applied to Physical Education"). The content of the lecture was carefully limited while the entertainment value was heightened. Dr. Fox was introduced as a world renowned authority in his specialty. The audience was asked to rate the speaker and offer

constructive comment. The speaker was highly rated by all the attendees with many asking for further information. Subsequently, Dr. Fox addressed several other professional groups with similar results. Naftulin et al [19].

Subsequent research attempted to demonstrate that the audience would be satisfied with the amount learned, as long as the lecture was high in expressiveness, and irrespective of the lecture's content. Ware and Williams, reported that student ratings were high under these circumstances even though student achievement was affected. In two subsequent studies Ware and Williams [20 & 21] demonstrated students were not sensitive to content unless the lecturer was low in expressiveness. Tending to identify a bias in student ratings due the seductiveness of the lecturer. In two other studies they attempted to sensitize students to the content of lectures and paid them for added learning. In neither of these studies did the students see through the Dr. Fox Effect. Williams and Ware [22 & 23].

The general premise of these studies: that lecturer expressiveness has a major impact on student ratings, has been confirmed by the studies of Perry et al [24], Meir Feldhusen [25], Frey [26], and Ramagli and Greenwood [27]. So a positive finding that can be identified with these studies is that an entertaining and expressive instructor will receive higher ratings than those who do not evidence these qualities. It is much more difficult to test the idea that content is not a necessary ingredient in high ratings. The various studies of Williams and Ware, Meir and Feldhusen, and Ramagli and Greenwood tend to support the idea while the study of Perry and his associates tends to show a different conclusion. Perry, Abrami, and Leventhal [28]. Abrami subsequently identified ten studies which used the expressiveness by content factorial design and subjected the studies to a meta-analysis employing omega-squared as an estimate of effect size. He found the expressiveness effect on student ratings to be quite large while expressiveness by content, and content effect were insignificant. However the content effect on achievement was large. Suggesting that instructor expressiveness was a major influence over student ratings but not on student achievement. Further he concluded that lecture content has a strong effect on achievement but not on student ratings. Based on these results he has suggested that the weak interaction on expressiveness provides evidence against the idea of educational seduction. Abrami [29]

However the Ramagli and Greenwood study found that: "students, when comparing lecture presentations varying in content and expressiveness are not able to identify differences in content level, or at the very least, become unaware of content level in a lecture presentation," they also concluded when looking at the same lecture given with and without high expressiveness: "low expressiveness of the first lecture influenced higher student ratings of the second lecture and that high expressiveness of the first lecture influenced lower student ratings of the second lecture." Without attempting to judge the results of any one researcher over another it should be evident that the "Dr. Fox Effect" does occur in the classroom but the appropriateness and strength of this influence is not clear.

Developments in Business Simulation & Experiential Exercises, Volume 12, 1985

Taken in a negative sense the Dr. Fox Effect, if completely true, suggests student ratings are fraudulent because the lecture lacking content is rated higher. If it is assumed that student ratings do what they are expected to do, then expressive teaching is rewarded by the students with higher ratings which in turn is rewarded by the administration and therefore must be the best approach.

Methodological Considerations

Viewed as an experiment to establish the effectiveness of any given teacher the student ratings appear to be quite limited. In using the word experiment it should be understood that the intent of the administration using ratings would be to measure the variables of relations under study. Although the rating technique is the most common technique used in research on teacher effectiveness, these instruments use numerical rating scales, check lists, and forced-choice ratings scales which presume the variables identified are the most important to effective teaching. Unfortunately there is not a widely accepted definition of good teaching nor are the variables under consideration clearly defined. Further, because of personal bias and a lack of training of the raters (faculty, administrators, and students), the results of using these instruments has always been poor and contradictory. Biddle and Ellena [30].

To satisfy some of the difficulties experienced with research in the past Returners suggested rating scales should satisfy the following criteria. Remmers [31]:

Objectivity. - The instrument should yield variable, reproducible data that is not a function of any peculiar characteristics of the rater.

Reliability - The rater should be accurate enough in his observations to allow results to be replicated under the same set of conditions.

Sensitivity - The rater should be able to make distinctions fine enough to communicate about the object of the investigation.

Validity - The categories on the rating scale should be relevant to some behavioral science construct satisfying the concepts of definition, construct, concurrent, and predictive validity.

Utility - The instrument should efficiently yield information relevant to contemporary theoretical and practical issues.

The issue of objectivity is especially important in behavioral observation. This stems from two potential sources of difficulty: the observer's powers of inference and the fact that the observer may be a part of the observational situation. Presumably the student is a competent observer who can exercise unbiased judgement. In addition it must be assumed that his role as a student will not influence his judgement of the teacher nor will his role as an observer influence the teacher's conduct of the class. Unfortunately research indicates that expected grades are a major influence over the rating given by the student. Bradenberg [32], Brown [33], Burton [34], Fagen [35], Hocking [36], Kau [37], Marsh [38], Powell [39]. Marsh has also found that interest in the course and workload/difficulty influence student raters. Trent [40] found that the values and attitudes of graduate students influenced the student ratings and Winne [41] found that student preference for a particular style of teaching influenced ratings. While Abrami [43] tried unsuccessfully

to prove that student or teacher characteristics did not influence ratings. Lavender [44] found lower ratings given to those with whom the rater experiences normative dissonance. Finally, Everett [45] found that students favored those professors emphasizing a lower level of cognitive material. Clearly objectivity is a serious problem in obtaining useful results from student ratings.

Reliability or dependability, or stability, or consistency, or predictability, or accuracy are assumed to be the same thing. High reliability does not signal good scientific results but to properly interpret results the test must be reliable. Reliability is usually controlled using instruments which are unambiguous and which are administered with clear instructions. Of course, the chance for error is larger if the instrument only has a few items. This result is associated with the higher probability for random error. More items increase the probability of an accurate measurement. Kerlinger [46]. Unfortunately, as student ratings are currently administered, there is only a need for a single item or if other items are listed only the student's rating of the instructor is considered in the analysis of results.

However, a number of researchers have concluded that the students are reliable Bligh [47], Centra [48], Murray [49], and Wood [50]. Significantly, both Centra and Wood suggest that faculty are less capable of rating their peers than the students. Leventhal [51] and Romney [52] indicate that ratings vary according to the courses taught and Wilson [53] found that the characteristics of a good Instructor varied across campus. There appears to be some evidence to support the idea that instructors teaching the same course from year to year will be rated about the same, the basis for the conclusion that student ratings are reliable. If true, it might be well to speculate about the psychological impact on the lowly rated teacher when ratings are administered quarter after quarter, assuming he is allowed to continue teaching.

Sensitivity of the rater assumes that the variable being measured is unambiguous. In this instance there should be a precise definition of effective teaching, so that the rater can look at a scale and perceive exactly what scale intervals mean. There is a need for a precise measure of what is good teaching as opposed to bad teaching. A 3.5 (on a scale of 5) should have a specific meaning while a 1.5 can be regarded as meaning something significantly different. Perhaps the most interesting finding with regard to sensitivity is that a non-major will rate a teacher differently than a student within the major Coles [54] or that the more knowledgeable the student, the higher the rating Haslett [55].

Validity is the critical dimension in any scientific investigation, and presents almost impossible difficulties for this type of measurement because it has so many dimensions. The initial question that must be addressed is quite simple yet difficult to answer with regard to student ratings: What is being measured? Or in this instance, what is effective teaching? Next, content validity might be considered or what bearing do the results of student ratings have on effective teaching? In a sense content validity relates to the representativeness of the sampling subjects' responses to a measurement instrument. Predictive validity is characterized by prediction to an outside criterion. The difficulty with predictive validation is that some criteria must exist which can be used as the basis for comparison. Finally construct validity may not be completely appropriate since it requires the verification of a scientific theory. Mere success in predicting

Developments in Business Simulation & Experiential Exercises, Volume 12, 1985

an outcome does not satisfy construct validation, although as noted earlier Remmers suggests the need to define the investigation in terms of "relevant behavioral science construct; if possible, the data should be covariant with some other, experimentally independent, index."

The question of the validity of student ratings has been addressed by researchers in two ways: comparing student ratings with ratings supplied by outside observers or faculty, and by measuring the achievement of students on examinations given in various sections in which the teacher has been rated by the students. Some research suggests that the student ratings confirm the judgement of trained outside observers Murray 156 & 57] while others find just the opposite to be true Greenwood and Renner 158), Greenwood [59]. Doyle and Crichton [60] found the correlation between student and faculty ratings to be very uneven, sometimes a good correlation and at other times poor. Gammell [61] found that the "aspects of instructor performance measured by student ratings may be quite separate from the factors that result in good teaching." This reluctance to accept student rating validity is echoed by others Benton and Scott 162], Bligh [63], Whitely and Doyle [64]. Murray [49] surveyed 30 studies using student achievement to validate student ratings and found that reported correlations between these two variables varied from -.75 to .70. He indicated the best studies showed validity.

Perhaps the research throwing the most doubt on the validity of student ratings is the finding that instructor reputation does affect student ratings McClelland [65], Perry, Niemi, and Jones [66], Perry, Abrami, Leventhal, Check [67]. This finding is related to the finding that student expectations play an important role in the way instructors are rated. So it might be said in summary that although the validity of student ratings has not been strictly demonstrated, there is some research favoring the ratings. Of course there is also a serious question about the applicability of the measurements especially when the same instrument is applied to various courses in various departments on the same campus. Administrators should be ready to specifically cite what is being measured. On many campuses it would be correct to indicate that the student is reporting on the satisfaction of expectations, O'Tuel [15], information that may not be vital in the administration of a department or school.

Utility suggests that an efficient means of measuring teacher effectiveness be adopted. Standardized student ratings appear to be an inexpensive way of learning what is going on in the classroom. But, possibly every research study covered in this paper indicates the need for additional sources of information in any faculty personnel action. This would suggest that student ratings are not as cost efficient as assumed. As a matter of fact, one researcher has suggested it would be too expensive to verify student ratings Centra ~46] while other researchers are so concerned about the ambivalence of validation research they question the use of student ratings for faculty personnel actions. Greenwood and Renner [58].

The contradictory conclusions of researchers suggest administrators should be cautious in their use of student evaluations. Administrators need to be well informed about limitations in the use of student ratings, while in fact, there is some evidence to indicate that administrators know little about the need for careful interpretation of student rating information. In this study, the review of the literature suggests three hypotheses which would merit further research and testing: 1) Student ratings are widely used by administrators for faculty personnel actions. 2) Student ratings constitute the major criteria in faculty personnel actions in institutions with a "teaching orientation." 3)

Administrators using student ratings in faculty personnel actions do not know if the rating system is either reliable or valid.

Pilot Study Conducted

The California State College System was selected for an exploratory study of the three hypotheses, because the Board of Trustees for the system has mandated the use of student evaluations as one element in the evaluation of teacher effectiveness. The student rating form used, the frequency with which ratings are taken, and the actual use of student rating information varies from campus to campus and within the individual schools on each campus.

The pilot study involved eleven open ended questions covering the specific use of student ratings and the administrator's understanding of the reliability and validity of the system they were using. Fourteen of those campuses with a management department were surveyed with the questions directed to the chairperson of the department. Table 1 contains a list of the survey questions and a summary of responses of Management Department chairpersons. Findings suggest little uniformity among these administrators either with regard to the importance of student ratings in their personnel actions or with regard to the administrator's understanding of the reliability and validity of the system being used. For example, in terms of the weight assigned to the student ratings as compared to other personnel evaluation criteria, one administrator estimated the ratings were weighted about 80% of the total, while another administrator estimated their importance at 15%. These results represent the spectrum of responses, a wider range in the importance accorded student ratings than anticipated. However, the findings otherwise confirmed the stated hypotheses: the student ratings were used by all the institutions in faculty personnel actions and the administrators appeared to have little objective information about the reliability and validity of the system they were using.

With respect to the hypothesis that student ratings constituted the major criteria in faculty evaluations, the results are mixed. At six of the campuses student evaluations are weighted 50% or more in faculty evaluations; at five campuses they are weighted 30% or less; at one campus they were used as a mandatory toll gate for further evaluation using other factors; and two administrators were unable to assign weights.

All but two of the administrators think student ratings are reliable, at least to some degree, but there is some confusion on the definition of reliability. All but four of the departments checked reliability across courses. Most perceive the evaluations to be valid but few had checked the content and predictive validity of the evaluation forms in use. Only two administrators reported a positive correlation between instructors receiving high ratings and the academic achievement of their students. The remainder did not know whether such correlation existed.

Clearly, a better understanding of rating system on the part of administrators is needed to help avoid some of the undesirable aspects of student rating systems and to help administrators avoid legal problems.

Conclusions

This paper has attempted to systematically review the literature covering the use of student ratings for faculty personnel actions. It can be concluded that

Developments in Business Simulation & Experiential Exercises, Volume 12, 1985

TABLE 1
Responses of Chairpersons of Management Departments in California State Universities and Colleges System

Universities:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Questions														
Are student ratings used for personnel decisions?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
What weight are they given (0 to 100)?	Don't know	30%	15%	17%	80%	Don't know - toll gate that must be passed	Varies	25%	25%	40-50% (Varies - IT or lecturer)	50%+	80%	50%	50%
Are the rating forms reliable?	Not totally	Yes--at ends of distrib.	Yes	Yes	Yes--reasonably	Yes--reasonably	No	Don't know	Yes	Yes	Yes	Yes	Yes # ends	50% reliable
How do you define reliability?	?	Stable, dependable	Consistent for an instructor	Don't know	Relation to actual performance	Stat. eval. of sample size	Lenient grades get better ratings	Is it statistically reliable?	Across courses	70%	Across courses	Reflects instr's capability?	Stability	Reflects instr. ability & student learning?
Is reliability checked across different courses for same instructors?	Yes	Yes	Yes--must be rated in all courses in one semester	Yes	Yes	Yes	No	No	No	Yes	Yes--all courses every semester	No	Yes	Yes
Are the ratings valid?	Partially	Probably	Yes	Qualified yes	Yes	Qualified no	Qualified yes	Qualified yes	Qualified yes	No	**Qualified yes	Yes	Qualified yes	Yes
How do you define validity?	Measures what is intended	Measures what is intended	Measures what it's supposed to	Performance compared to others vs ratings	Measures what it's supposed to	To check for halo effect	Measures what is intended?	Measures what it's intended to?	Measures what is intended?	Judgment	?	Same as reliability	Measures what is intended?	How much students enjoy course
Has the content validity of your ratings forms been checked?	Yes	Only one question	Yes--two years ago--revised	Yes--Univ. wide study--forms revised	Continuously	No	No	Yes	No	No	Yes	Don't know	No	No
Has the predictive validity of your rating forms been checked?	No	No	No	No	No	No	No	Yes	No	No	No	No	Not statistically	No
Do students of instructors with high ratings have higher academic achievement?	Don't know	Don't know--probably not	Don't know	Don't know	Don't know	Not tested	Yes	Don't know	Don't know	Don't know	Don't know	Not sure	Don't know	Yes
Has this been verified?	No	No	No	No	No	No	NO	No	No	No	No	No	No	Yes

*California State University: Sacramento, Los Angeles, Chico, Dominguez Hills, Fresno, Fullerton, Hayward, Northridge, Cal. Polytechnic State Univ.-San Luis Obispo; Cal. State Polytechnic Univ.-Pomona; San Francisco State Univ.; San Jose State Univ.; Cal. State College-Stanislaus; Humboldt State Univ. (This listing does not correspond with numbers at top of table.)

**[In terms of student satisfaction, but not in terms of what they're supposed to measure.]

student ratings are strongly supported by some researchers but that even these researchers suggest student ratings be adopted as only part of the appraisal of college teaching performance.

Used correctly, that is with due respect to their limitations, student ratings can offer useful information. Otherwise, the ratings probably should not be taken too seriously. As the results of the pilot study suggest, administrators should be more concerned about the reliability and validity of the student rating information.

A major problem associated with the use of student rating systems which has not been explored is their effect on Institutional standards. The research on the effect of expected grades on student ratings would suggest that faculty may become lenient in grading standards under the assumption that higher student ratings will result. This and other potential problems affecting institutional standards, such as: the amount of material covered in any given course, the quality of textbooks adopted, and the rigor of courses should be researched to determine the effect of student rating systems.

Student ratings give the students a feedback channel and demonstrate to the public that faculty are interested in doing a good job in the classroom. On the other hand, these ratings also give administrators a seemingly remote control over classroom activity that may not always accomplish what is intended.

FOOTNOTES

- [1] J.M. Stephens, The Process of Schooling: A Psychological Examination, (New York, Holt, Rinehart and Winston, Inc., 1967).
- [2] J.C. Biddle, "Annotated Bibliography of ERIC Reports Concerning Student Evaluations of Faculty Performance for Years 1974 thru 1978," ERIC Report ED: 398 832.
- [3] Brondon Beatty and Herbert W. Marsh, "Students' Evaluations of Instructional Effectiveness," ERIC Report ED: 141 413, May 1974.
- [4] Warren Bechtolt and Robert Hawkins, "Method Evaluates Student Evaluation of College Courses," Journalism Educator, 32, 4, January 1978.
- [5] W.J. McKeachie, Y. Lin and W. Mann, "Student Ratings of Teacher Effectiveness: Validity Studies," American Educational Research Journal, 8, 1971.
- [6] Herbert W. Marsh and J.U. Overall, "Validity of Students' Evaluations of Teaching: A Comparison with Instructor Self Evaluations by Teaching Assistants, Undergraduate Faculty and Graduate Faculty," ERIC ED 177 205, April 1979.
- [7] Raymond P. Perry, "Educational Seduction: Implications for the Evaluation and Improvement of Teaching," ERIC: ED 233 638.

Developments in Business Simulation & Experiential Exercises, Volume 12, 1985

- [8] A.M. Sullivan and G.R. Skates, "Validity of Student Evaluation of Teaching and the Characteristics of Successful Instructors," Journal of Educational Psychology, 66, 1974.
- [9] N. Bennington, "Student Evaluation of Teaching in Management and Business Studies," ERIC; ED 155 954,
- [10] Trudy Bers, "An Analysis of the Student Evaluation Form at Oakton Community College," ERIC: ED 140 877.
- [11] Donald Bligh, "The Reliability Yet Invalidity of Students' Judgement of Teaching," ERIC: ED 177 932.
- [12] Mary G. Chisholm, "Provocative Opinion: Student Evaluation: The Red Herring of the Decade," Journal of Chemical Education, 54 1, 1977.
- [13] Eli Cohen, "Issues Confounding the Measurement of Teaching at CSUS," ERIC: ED 238 950.
- [14] Raymond Moody, "Student Achievement and Student Evaluations of Teaching in Spanish," Modern Language Journal, 60, 8, December 1976.
- [15] Frances S. O'Tuel, "Student Ratings for an Ideal Professor," ERIC: ED 192 648.
- [16] M. Rodin and B. Rodin, "Student Evaluations of Teachers," Science, 177, 1972.
- [17] J.E. Boyd and E.F. Schietinger, Faculty Evaluation Procedures in Southern Colleges and Universities, (Atlanta, Southern Regional Educational Board, 1976).
- [18] E. Goffman, The Presentation of Self in Everyday Life, (New York, Doubleday Books, 1959).
- [19] D.H. Naftulin, J.E. Ware and F.A. Donnelly, "The Dr. Fox Lecture: A Paradigm of Educational Seduction," Journal of Medical Education, 18, 1973.
- [20] J.E. Ware and R.G. Williams, "The Dr. Fox Effect: A Study of Lecturer Effectiveness and Ratings of Instruction," Journal of Medical Education, 50, 1975.
- [21] J.E. Ware and R.G. Williams, "Discriminant Function Analysis of Student Rating as a Means of Identifying Lecturers Who Differ in Enthusiasm or Information Giving," Education and Psychological Measurement, 37, 1977.
- [22] R.G. Williams and J.E. Ware, "Validity of Student Ratings of Instruction Under Different Incentive Conditions: A Further Study of the Dr. Fox Effect," Journal of Educational Psychology, 68 (1), 1976.
- [23] R.G. Williams and J.E. Ware, "An Extended Visit With Dr. Fox: Validity of Student Ratings of Instruction After Repeated Exposures to a Lecture," American Educational Research Journal, 14(4), 1977.
- [24] R.P. Perry, P.C. Abrami and L. Leventhal, "Educational Seduction: The Effect of Instructor Expressiveness and Lecture Content on Student Achievement," Journal of Educational Psychology, 71, 1979.
- [25] R.S. Meir and J.F. Feldhusen, "Another Look at Dr. Fox: Effect of Stated Purpose of Evaluation, Lecturer Expressiveness, and Density of Lecture Content on Student Ratings," Journal of Educational Psychology, 71, 1979.
- [26] P.W. Frey, "The Dr. Fox Effect and Its Implications," Instructional Evaluation, 3(2), 1979.
- [27] H.J. Ramagali and G.E. Greenwood, "Another Assessment of the Dr. Fox Effect: Paired Lecture Comparisons of Expressiveness and Content," Journal of Educational Psychology, 1981.
- [28] See Perry et. al. p. 116.
- [29] P.C. Abrami, Educational Seduction, Unpublished manuscript, Concordia University, Montreal, Canada, 1981.
- [30] B.J. Biddle and W.J. Ellena (Eds), Contemporary Research on Teacher Effectiveness, (New York, Holt, Rinehart and Winston, 1964) p. 26.
- [31] H.H. Remmers, "Rating Methods in Research on Teaching," in N.L. Gage, Handbook of Research on Teaching, (Chicago, Rand McNally, 1963) p. 330.
- [32] Dale C. Brandenburg, et. al., "Student Ratings of Instruction: Validity and Normative Interpretations," Research in Higher Education, 7, 1, 1977.
- [33] David Lila Brown, "Faculty Ratings and Student Grades: A University-Wide Regression Analysis," Journal of Educational Psychology, 68, 5, October 1976.
- [34] Gene Burton, et. al., "Effects of Contract Grading on Student Evaluations of Teaching," Educational Horizons, 55, 3, Spring 1977.
- [35] Nancy Magnusson Fagan, et. al., "Self Esteem, Expectation and Attribution of Responsibility for Academic Performance," ERIC: ED 239 167, 1983.
- [36] Joan M. Hocking, "College Students' Evaluations of Faculty are Directly Related to Course Interest and Grade Expectation," College Student Journal, 10, 4, Winter 1976.
- [37] James B. Kau and Paul H. Rubin, "Measurement Techniques, Grades and Ratings of Instructors," Journal of Economic Education, 8, 1, February 1976.
- [38] Herbert W. Marsh, "Students' Evaluations of Instructional Effectiveness: Relationship to Student, Course, and Instructor Characteristics," ERIC: ED 155 217.
- [39] Robert W. Powell, "Grades, Learning, and Student Evaluation of Instruction," Research in Higher Education, 7, 3, 1977.
- [40] Curtis Trent and James F. Johnson, "The Influence of Students' Values and Educational Attitudes on Their Evaluation of Faculty," Research in Higher Education, 7, 2, 1977.
- [41] P.H. Winne, "Aptitude - Treatment Interactions in an Experiment on Teacher Effectiveness," American Educational Research Journal, 14(4), Fall 1977.

Developments in Business Simulation & Experiential Exercises, Volume 12, 1985

- [42] Philip C. Abrami, et. al., "The Generalizability of Student Ratings of Instruction," ERIC: ED 139 832.
- [43] Abraham D. Lavender, "Dissonance as a Factor in College Student Evaluation of Faculty," College Student Journal, 11, 2, Summer 1977.
- [44] Les Leventhal, et. al., "Bogus Evidence for the Validity of Student Ratings," ERIC: ED 150 510.
- [45] Michael D. Everett, "Student Evaluations of Teaching and the Cognitive Level of Economics Courses," Journal of Economic Education, 8, 2, Spring 1977.
- [46] Fred N. Kerlinger, Foundations of Behavioral Research, (New York, Holt, Rinehart and Winston, 1964) p. 442.
- [47] J.A. Centra, "Reliability of Student Instructional Report Items," SIR Report #3, Princeton, New Jersey: Educational Testing Service 1973.
- [48] J.A. Centra, "Colleagues as Raters of Classroom Instruction," Journal of Higher Education, 46, 1975.
- [49] Harry G. Murray, "Student Evaluation of University Teaching: Uses and Abuses," ERIC: ED 189-920.
- [50] Peter H. Wood, "Student and Peer Ratings of College Teaching and Peer Ratings of Research and Service: Four Years of Departmental Evaluation," ERIC: ED 155 218.
- [51] Les Leventhal, P.C. Abrami, R.P. Perry and L.J. Breen, "Section-Selection in Multi-Section Courses: Implications for the Validation and Use of Teacher Rating Forms," Educational and Psychological Measurement, 35, 1975.
- [52] David Romney, "Course Effect vs. Teacher Effect on Student's Ratings of Teacher Competence," Research in Higher Education, 5, 4, 1976.
- [53] Pamela Wilson and Terry C. Wilson, "What Factors Contribute to Better Instruction? Business Students vs. Other Colleges," ERIC: ED 153 009.
- [54] William H. Coles, "College Student Perceptions, 1976 Freshman and Seniors. Expectations and Perceptions of Faculty," ERIC: ED 152 206.
- [55] Betty J. Haslett, "The Influence of Student Knowledgeability on Student Rating of Instruction," Communication Education, 26, 1, 1976.
- [56] Harry G. Murray, "The Validity of Student Ratings of Faculty Teaching Ability," Paper presented at the Canadian Psychological Association Meeting, 1972.
- [57] Harry G. Murray, "How Do Good Teachers Teach? An Observational Study of the Classroom Teaching Behaviors of Social Science Professors Receiving Low, Medium, and High Teacher Ratings," Ontario Universities Program for Instructional Development Newsletter, February 1977.
- [58] G.E. Greenwood and R.R. Renner, "Student Ratings of College Teaching: Their Validity and Use in Administrative Decision-Making," Science Education, 59(4), 1975.
- [59] Gordon Greenwood, "Student Evaluation of College Teaching Behaviors," ERIC: ED 134 868.
- [60] K.O. Doyle and Crichton, "Student, Peer, and Self Evaluation of College Instructors," Journal of Educational Psychology, 70, 1978.
- [61] Linda Gammell, "Student Ratings, Other Measures Help Assess Teaching Impact," ERIC: ED 143 276 May 1977.
- [62] Sidney E. Benton and Owen Scott, "A Comparison of the Criterion Validity of Student Response Inventories for Appraising Instruction," ERIC: ED 128 397, April 1976.
- [63] See Bligh, p. 157.
- [64] Susan Whitely and Kenneth Doyle, "Implicit Theories" in American Educational Research Journal, 13, 4, Fall 1976.
- [65] J. McClelland, "The Effect of Student Evaluations of College Instruction Upon Subsequent Evaluations," California Journal of Educational Research, 2, 1970.
- [66] R.P. Perry, R.R. Niemi, and K. Jones, "Effect of Prior Teaching Evaluations and Lecture Presentation on Ratings of Teacher Performance," Journal of Educational Psychology, 66, 1974.
- [67] R.P. Perry, R.R. Niemi, and J. Check, "Instructor Reputation: An Expectancy Relationship Involving Student Ratings and Achievement," Journal of Educational Psychology, 71, 1979.