J. INDIVIDUAL EMPLOYMENT RIGHTS, Vol. 2(1) 73-86, 1993-94

STUDENT EVALUATION OF INSTRUCTION: OBJECTIVE EVIDENCE AND DECISION MAKING

DAVID A. DILTS

HEDAYEH SAMAVATI

MASHALAH RAHNAMA MOGHADAM

LAWRENCE J. HABER

Indiana University at Fort Wayne

ABSTRACT

This article examines student attitudes as a possible source of contamination of evidence gathered from student evaluation of instruction questionnaires. A student attitude survey instrument was administered simultaneously with a student evaluation of instruction instrument for undergraduate students and for graduate students at three different midwestern universities. The results demonstrate that 13 percent of the variation in undergraduate student appraisals of instruction is explained by student attitudes such as personal dislike for an instructor, racism or sexism. Graduate students' appraisals of instructional effectiveness are not correlated with inappropriate attitudes. These findings suggest that decision making by undergraduate students concerning the quality of instruction they received are significantly affected by inappropriate attitudes, thereby creating suspicion concerning the objectivity of undergraduate student evaluators. In turn, academic decision makers must exercise care in interpreting evidence gathered using student evaluation questionnaires administered in undergraduate classes. There is no evidence that graduate students' evaluation should be regarded by academic administrators with equal suspicion.

It has been conjectured that decisions in academic personnel matters may not be guided by objective evidence [1]. In fact, Schroeder concluded that little organized effort has been made to apply standard management science techniques to academic decision making [2]. Since the publication of Schroeder's work,

doi: 10.2190/TG7W-MK7A-KNLB-2WAD
http://baywood.com

several other studies have indicated that there will likely be more objective criteria used to evaluate faculty performance [3, 4]. The appropriate modeling of academic performance critically depends on the gathering of reliable, objective evidence. Further, the recent changes in the American Assembly of Collegiate Schools of Business (AACSB) accreditation guidelines focus more attention on the quality of classroom instruction. Thus, the changes provide an impetus to serious efforts to improve teaching [5, 6]. To assess whether efforts aimed at improving teaching have been effective, more careful examination of the evaluation of instruction is necessary [7]. The use of instructional evaluation questionnaires has gained wide acceptance as one of the primary assessment techniques in faculty performance evaluation models [8]. Because student questionnaires have become a common (and too often exclusive) method of evaluating instruction, the decision making of students in responding to these questionnaires requires systematic analysis.

The use of teacher evaluation questionnaires is a controversial issue in higher education. The student activism of the 1960s led to increasing acceptance of these evaluations both for the development of instructional skills of faculty and for performance assessment. Questions of validity and reliability of the student evaluation of instruction have led to extensive research concerning the subject [8, 9]. Proponents and opponents of the use of student evaluations of instruction have both found support in published studies. The conflicting results of studies concerning the validity and reliability issues in student evaluation of instruction have fueled the continuing debate. In fact, Marlin and Niss observed that the essence of the controversy can be reduced to what they called the Marlin/Niss Principle:

> Teachers who have received high student evaluations in the past will find them to be valid measures of good teaching. Teachers who have received low student evaluations in the past will find them to be laughably insignificant [10, p. 25].

Aigner and Thum reported that as much as two-thirds of the variance in student ratings of professors is determined by the characteristics of the individual professors [11]. Professorial traits such as enthusiasm, organization, and rapport were found to significantly affect student responses to questionnaire items. The characteristics that Aigner and Thum tested were associated with both the cognitive and affective domains of instruction [11]. Cashin, among others, argued that the quality of the instruction received determines a significant proportion of the student responses [12]. The results of these studies indicate that a large proportion of student decision making is consistent with the goals of objective instructional assessment. What is left unanswered is whether some significant proportion of student decision making is clouded by inappropriate bases, for example, such as racism or sexism.

While student responses to teacher evaluation questionnaires may be determined by factors that legitimately determine quality of instruction, the unalterable fact is that the responses are filtered through the students' perceptions of that instruction. Consequently, student attitudes and perceptions may influence students' responses to teacher evaluation questionnaires. The purpose of this study is to determine whether there is any evidence to support the conclusion that students' attitudes affect their evaluation questionnaire responses.

DATA

The data were gathered using paired questionnaires. The teacher evaluation form used in the School of Business and Management Sciences at Indiana University-Purdue University at Fort Wayne (IPFW) was administered to students together with the survey of students' attitudes toward instructional evaluation. The evaluation questionnaire was the first page, and the attitude survey was the second page of the two-page instrument. The teaching evaluation form contained ten items, and the attitude survey contained eleven items. Table 1 presents the items and associated variable names from the student evaluation questionnaire. Table 2 presents the items and associated variable names from the attitude survey questionnaire.

The student evaluation of teaching questionnaire was constructed by a committee of faculty members from the School of Business and Management Sciences at IPFW. The questionnaire is unremarkable, save that it is relatively short. This fact permits efficiency in data gathering at the expense of sampling only certain aspects of teaching quality.

The attitude survey was constructed from a review of the literature concerning possible biases observed in student evaluations of instruction (e.g., [13, 14]). From this survey eleven possible biases were identified (see Table 2). The attitude survey was field tested using a sample of thirty-four nonbusiness students from the IPFW campus to ascertain precisely how respondents interpreted the items. Three questions were identified by these students as having multiple interpretations. The three items were rewritten and again the students were quizzed concerning their meaning. Once the thirty-four students agreed on the intended interpretation of all the items, the questionnaire was administered to a larger sample. The sample consisted of 150 business students from three undergraduate sections at IPFW in the spring semester of 1989. The purpose of surveying the larger sample was to determine whether variation in the responses could be observed. Variations in the responses from the spring 1989 field test were observed, and the study was subsequently undertaken using the paired questionnaires. The data obtained from the student attitude survey were used as categorical explanatory variables in the regression analysis.

The questionnaires asked students to respond from strongly agree to strongly disagree on a five-point Likert scale on each of the items. The questionnaires were

Table 1. Teacher Evaluation Questionnaire

Questionnaire Item	
1. [INFORMED]	The instructor appears to be *well informed* on this subject.
2. [EXPLAIN]	The instructor's explanations of course content are clear and to the point.
3. [COMMO]	The instructor possesses an ability to communicate course expectations and goals effectively.
4. [HELPFUL]	The instructor demonstrates a helpful and concerned attitude toward students.
5. [EXAMS]	The content of examinations appears to be both appropriate and applicable to the objectives of this course.
6. [TEXT]	The course materials and textbook readings are useful and contribute to an understanding of course content.
7. [ORGANIZED]	The course content is effectively organized in terms of course goals and objectives.
8. [METHODS]	The instructor uses teaching methods helpful for learning the content of this course.
9. [COURSE]	Overall, this is a valuable and informative course.
10. [INSTRUCTR]	Overall, this is a capable instructor.

administered to 302 undergraduate students in seven sections and seventy-one Master of Business Administration (MBA) students in four sections at three universities: Indiana-Purdue University at Fort Wayne, Winona State University, and Cleveland State University. The instructors in the undergraduate sections included three foreign-born professors, two of whom were female. The remaining instructors were white males. The instructors for the graduate sections included one foreign-born female in two sections, while the other instructors were white males.

The descriptive statistics for undergraduate and graduate student responses to the items on both the teacher evaluation form and the attitude survey are included in Tables 3 and 4 respectively.

The statistics presented in Table 3 indicate that the average rating given by graduate students is higher than that given by the undergraduates on each item in the teacher evaluation questionnaire. The variability of the graduate student responses was slightly less than the variability of the undergraduate responses in most cases. The descriptive statistics in Table 4 indicate that, when compared to the graduate students, the undergraduates reported that they were influenced more (on average) by attitudes not strictly related to the quality of instruction. Thus,

Table 2. Survey of Student Attitudes Toward Evaluation of Instruction

Questionnaire Item	
1. [RACE]	My responses to an evaluation questionnaire have been influenced by my instructor's race, sex, or ethnic background.
2. [GRADE]	My responses to an evaluation questionnaire have been influenced by the grade I expected to receive in the course.
3. [GET EVEN]	I have used the evaluation questionnaire to "get even" with professors I don't like.
4. [ENTERTAIN]	I give instructors who are entertaining higher marks on evaluation questionnaires.
5. [VALUES]	I give instructors who share my values and views higher marks on evaluation questionnaires.
6. [REQUIRED]	I give instructors who teach required courses lower marks than those who teach elective courses.
7. [SERIOUS]	I take the student evaluation questionnaires seriously.
8. [JUDGE]	Students are good judges of whether they received quality instruction.
9. [EXPERT]	Students are good judges of whether professors know their field of expertise.
10. [FELLOW]	My fellow students do not take the evaluation process very seriously.
11. [IMPORTANT]	The student evaluation questionnaire is a very important element to assure quality instruction.

the mean responses to RACE, GRADE, GETEVEN, ENTERTAIN, VALUES and REQUIRED are all higher for the undergraduates than for the graduate students. In addition, undergraduates reported that, on average, they took the teacher evaluation process less seriously than did the graduate students.

There is, however, another aspect of the data that requires discussion. The authors made a concerted effort to ask only those instructors with approximately equal instructional abilities, and who were "good" teachers, to participate in this study. This procedure was followed to eliminate the possible interdependence of variation in professors' quality of instruction and stated student attitudes. To the extent that these controls are effective in limiting variations in instructional performance, there may be no significant variation in the data due to instructional quality. This design, however, is based only on historic evidence and perceptions of colleagues as to the quality of instructors' teaching. There is no scientific

Table 3. Descriptive Statistics for the Teacher Evaluation Items

	Undergraduate		Graduate	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
INFORMED	4.47	0.64	4.52	0.68
EXPLAIN	3.93	0.99	4.33	0.70
COMMO	3.88	0.96	4.32	0.70
HELPFUL	4.21	0.80	4.35	0.78
EXAM	4.10	0.77	4.45	0.79
TEXT	3.93	1.01	4.03	0.96
ORGANIZED	3.93	0.84	4.17	0.78
METHODS	3.84	0.95	4.13	0.91
COURSE	4.05	0.87	4.32	0.77
INSTRUCTR	4.26	0.85	4.48	0.72

Table 4. Descriptive Statistics for the Attitude Variables

	Undergraduate		Graduate	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
RACE	1.59	0.85	1.2	0.51
GRADE	2.44	1.13	1.83	1.01
GETEVEN	1.87	0.97	1.48	0.81
ENTERTAIN	3.37	1.06	2.81	1.29
VALUES	3.03	1.00	2.12	0.96
REQUIRED	1.82	0.74	1.47	0.70
SERIOUS	3.81	0.84	3.92	1.00
JUDGE	4.05	0.77	3.87	0.98
EXPERT	3.45	0.89	3.43	0.93
FELLOW	2.88	0.83	2.83	0.83
IMPORTANT	3.55	1.02	3.28	1.22

evidence to demonstrate conclusively when these instructors are equivalent save in gender, place of national origin, or religion. All the authors can claim is that a subjective evaluation approximating a promotion and tenure framework was used to select teachers who the authors believed were equivalent in instructional competence. In addition, to see whether there were significant differences in the quality of instruction that could be subjectively identified, student comments concerning communication abilities, knowledge of the subject and so forth, were elicited. No evidence from the student comments suggested a significant qualitative distinction between any of the instructors in the sample.

METHOD

The hypothesized relation between students' attitudes and their appraisal of instructional performance is presented in equation (1):

$$(1)\ \text{SET} = \alpha + \beta_1\,\text{RACE} + \beta_2\,\text{GRADE} + \beta_3\,\text{GETEVEN} + \beta_4\,\text{ENTERTAIN} + \beta_5\,\text{VALUES} + \beta_6\,\text{REQUIRED} + \beta_7\,\text{SERIOUS} + \beta_8\,\text{JUDGE} + \beta_9\,\text{EXPERT} + \beta_{10}\,\text{FELLOW} + \beta_{11}\,\text{IMPORTANT} + \varepsilon$$

The dependent variable, SET, is the average of the items from the student evaluation form presented in Table 1. Equation 2 presents SET:

$$(2)\ \text{SET} = (\text{INFORMED} + \text{EXPLAIN} + \text{COMMO} + \text{HELPFUL} + \text{EXAMS} + \text{TEXT} + \text{ORGANIZED} + \text{METHODS} + \text{COURSE} + \text{INSTRUCTR})/10$$

Equation 2 shows that the responses to the ten items on the student evaluation of instruction were averaged for each student to obtain a general, approximately continuous measure (within the range of one through five in gradations of one tenth) of the students' perceptions of the quality of instruction. (Note that the questionnaire items are constructed so that five is the most favorable response on all items.) Although one may argue that both COURSE and INSTRUCTR, the overall beauty measures, are just composite measures of the other eight items, the low correlation between the two overall measures and the other items suggests that they contain independent information about the quality of instruction. Thus, they are included in SET.

Ordinary least squares (OLS) was used to estimate equation (1) for undergraduate students' responses and for graduate students' responses. Because the average evaluation for each student is a near continuous variable, albeit within a range, ordinary least squares (OLS) can be applied.

EMPIRICAL RESULTS

Table 5 presents the results of the estimated equations for both graduate courses and undergraduate courses.

There is no evidence of heteroscedasticity or multicollinearity in the estimated equations. The highest correlation between the independent variables was about .30. Partitioning the data by institution gave consistent parameter estimates; therefore, the coefficients are stable. Accordingly, it appears that the data does not violate the ordinary least squares assumptions.

Table 5. Regression Results

Variables	Undergraduates	Graduates
INTERCEPT	4.243	4.750
	(11.290)*	(5.046)*
RACE	–0.096	–0.226
	(–1.914)**	(–1.392)
GRADE	0.009	0.077
	(0.218)	(0.811)
GETEVEN	–0.107	–0.005
	(–2.432)*	(–0.048)
ENTERTAIN	0.089	–0.049
	(2.152)*	(–0.635)
VALUES	0.084	0.057
	(1.807)**	(0.562)
REQUIRED	–0.110	–0.081
	(–1.895)**	(–0.673)
SERIOUS	0.119	0.055
	(2.395)*	(0.537)
JUDGE	–0.121	–0.094
	(–2.375)*	(–0.958)
EXPERT	–0.0917	0.052
	(–2.121)*	(0.498)
FELLOW	–0.011	–0.135
	(–0.244)	(–1.053)
IMPORTANT	0.050	0.056
	(1.276)	(0.735)
ADJUSTED R^2	0.128	0.000
F-STATISTIC	4.882*	0.950

Note: Numbers in parentheses indicate the t-values.
*Significant at the 0.05 level.
**Significant at the 0.10 level.

Table 5 shows that none of the variance in the graduate students' evaluations of teacher performance was explained by the items from the attitude survey. None of the coefficients for the attitude variables was significant for the graduate student equation. The estimated intercept term approximated the mean of the average evaluation for the graduate students; at about 4.75 on a five-point scale.

The undergraduate equation shows that student attitudes influence their ratings of instruction. The attitude variables explain about 13 percent of the variation in instructor evaluation. Eight of the eleven coefficients are significant, with only GRADE, FELLOW, and IMPORT being insignificant. The estimated intercept term for the undergraduate equation is 4.24, which is slightly higher than the mean of observed evaluations.

The estimated coefficients for RACE, GETEVEN, REQUIRED, JUDGE, and EXPERT are negative. ENTERTAIN, VALUES, and SERIOUS have positive estimated coefficients. The evidence supports the hypothesis that student attitudes negatively (positively) affect performance evaluations when students (do not) consider race, gender, or ethnic background of instructors, (do not) feel free to use the evaluation for personal vendettas, the course is (not) required, they (do not) believe that they are good judges of instruction or of the professors' expertise in the subject. The performance evaluation is positively (negatively) affected when students consider a professor (not) to be entertaining, (not) to share their values, or students (do not) take the evaluation process seriously.

DISCUSSION

The equation for the graduate students shows that the attitudes tested here do not affect their appraisals of instructional performance. All of the MBA candidates already possessed a bachelor's degree. It may be that the experience gained as an undergraduate results in a level of personal and academic maturity that mitigates the influence of attitudes on instructional performance appraisal. If this is the case, then only instructor characteristics and the quality of instruction will explain the variations in student evaluation of instruction for graduate courses.

The attitude survey was constructed using information published about undergraduate evaluation of instruction and was field tested using undergraduate students only. It may also be possible that graduate students have attitudes that are significantly different from those observed in undergraduate students. Consequently, the attitude questionnaire used in this study may not be applicable to graduate students. There may be other attitudes common to graduate students that will explain some significant proportion of variation in their evaluations of instruction. The evidence reported here does not lend any insight into which explanation for the zero correlation is most plausible. The remainder of the discussion focuses exclusively on the undergraduate equation.

The results show that students' attitudes have a significant effect on undergraduate student evaluations of instruction. RACE is negatively correlated with student appraisal of instruction. In other words, those students who agreed that their responses were influenced by race, gender or ethnic background downrated instructor performance, *ceteris paribus*. On the other hand, those students who disagreed with this statement gave more lenient evaluations.

An instructor's ability to communicate may be affected by race or ethnic background. To the extent that communication ability is correlated with race or ethnic background, this could be a measure of a legitimate appraisal issue. The correlation coefficient for RACE and COMMO for the undergraduate students, however, is –.087 which is insignificant. Thus, the negative coefficient for RACE is not the result of any communication difficulty that a foreign-born faculty member may have. For the sample of students, the remaining attitude variables are not significantly correlated with RACE. The evidence is consistent with some modicum of prejudice for those who downrate instructors' performance; other explanations may be possible, but unfortunately, they escape the authors. For those students unconcerned about race, gender, or ethnic background of the instructor, the evidence suggests more positive appraisals in general.

The coefficient for GETEVEN is perhaps more telling and is consistent with the findings reported for RACE. GETEVEN is negatively associated with the performance appraisal. This variable is a direct measure of student propensity to use the evaluation process for other than its intended purpose. Rather than a measure of instructional effectiveness, students perceived the evaluation process as a method of punishing those instructors they did not "like." The evidence also shows that students who disagreed with this statement tended to give more favorable appraisals of instruction.

It is possible that students could interpret GETEVEN as asking them to assess whether they liked the instructor's performance. The field tests, however, clearly show that all thirty-four students asked to interpret the question believed they were being asked to respond according to whether they liked the instructor personally.

The coefficient for REQUIRED is negative. This result is consistent with the results reported in several other studies of teaching evaluations [15]. The coefficient for this variable shows that a significant number of students resented being forced to take certain courses and downrated the instructors of those courses accordingly. A significant number of students did not resent being forced to take specific classes. These students did not exhibit a propensity to downrate instructional performance.

Both JUDGE and EXPERT have negative coefficients. In other words, a significant number of students downrated instructors when they believed they could effectively judge both the quality of the instruction they received and the expertise the instructor possessed in the subject matter. Alternatively, several students rated

instructors higher when they did not believe they could judge the quality of instruction or the instructor's expertise.

There is a propensity to leniency in rating the performance of employees with whom the rater is personally familiar or in rating those for whom a rater recognizes that he or she is not competent to perform the evaluation. The halo effect explains student responses that rank instructors highly when they did not believe they were capable of judging instructional performance.

That students believe they are good judges of instruction is not difficult to accept. That students are actually good judges of the expertise of faculty members is more difficult to allow. Undergraduates, particularly in introductory courses, do not have sufficient background or experience to judge the professional training of their instructors. In this regard, the graduate students surveyed were, on average, slightly less willing to judge themselves competent to assess either the instructional or the intellectual expertise of their instructors (Table 4).

The coefficient for ENTERTAIN is positive. Students who agreed that they give higher marks to entertaining instructors gave higher evaluations in general. Those who disagreed were more critical. More respondents fell in the former than in the latter category. This evidence shows that a significant number of students rated instructors higher when they valued entertainment in classroom instruction. On the other hand, those respondents who did not value entertainment downrated instructors. Those students who looked to be entertained were more lenient on student evaluation questionnaires, while those who did not care for entertainment tended to be more critical.

VALUES has a positive coefficient. Students who reported that they gave instructors higher evaluations when the instructor shared their values, in fact, gave higher performance appraisals. Those who disagreed with the statement gave lower evaluations. In responding to this question, the sample was about evenly split for undergraduate students. As Seigfried and Fels argued, education requires changes in students' values and changes in students' attitudes toward the subject and toward education in general [16]. The impact of these changes in values and in attitudes on subsequent behaviors are critical outputs of education. If students use differences in values and views as a basis for inflicting poor evaluations or if they use congruence in values and views to reward professors with higher evaluations, then the student evaluation process poses a potentially damaging constraint on education.

SERIOUS has a positive coefficient. Students who reported that they took the evaluation seriously rated instructors more favorably. Those students who reported that they do not take the process seriously tended to downrate instructors' performances. SERIOUS may be interpreted as an empowerment variable. In other words, the evaluation process may be viewed by students as an opportunity to praise effective instruction, but students do not believe that complaints of instructional deficiency will be well-received. It could be argued that those who

did not take the process seriously downrated instructors' performances, which suggests that these students did not believe their complaints or suggestions for instructional improvement are of importance to professors or administrators. It is also possible that students observed very little weight was placed on their opinions by those charged with making personnel decisions and that they may be expressing dissatisfaction with the evaluation process as well as instruction.

IMPLICATIONS FOR EVALUATION PROCESS AND CONCLUSIONS

There are several implications of the reported results for decision making in both instructional assessment and personnel policy. The attitudes of undergraduate students are significantly correlated with the results of the student evaluation of instruction. There is no evidence that graduate student evaluations are influenced by the attitudes examined in this study.

The results suggest that certain characteristic attitudes of undergraduate students are associated with favorable instructor appraisals, while others are associated with more harsh decisions concerning instructional quality. Attitudinal profiles of student evaluators can be constructed from the evidence presented in this study. Downrating of instructors because of racism, sexism, personal vendetta, disregard for differences in values and a disregard for the evaluation process do not serve the educational objectives of any university. On the other hand, the favorable appraisal of performance of instructors associated with more liberal student attitudes is inconsistent with the purpose of performance appraisal. Upwardly biased evaluations fail to convey accurate information that may be necessary to improve classroom instruction.

The evidence demonstrates that 13 percent of the variation in student evaluations is explained by student attitudes. The attitudes used as the independent variables in this study are those attitudes that people often hesitate to admit. That the respondents admitted to patently unacceptable motives in answering the questionnaires is surprising because of their natural propensity to avoid giving responses that are socially repugnant (social acceptance response bias). Because of this bias, the 13 percent of the variation explained may underestimate the magnitude of the influence of these attitudes on students' decisions concerning instructional assessment. Thus, depending on the attitudes prevalent in a particular class, an instructor's evaluations could be dramatically higher or lower than performance would merit. The results of this inquiry support the conclusion that caution must be exercised in using student evaluations in personnel decisions.

Personnel policies must account for the possibility that student evaluations are correlated with contaminating student attitudes. Some students' attitudes may be antithetical to educational objectives. Policy makers must recognize that student attitudes can and do cause higher or lower evaluations than an instructor would otherwise earn. It is plausible that there will be systematic bias in evaluation data

for specific classes. The bias will result from the attitudes of a significant number of students in the class and could result in significant differences in performance appraisals.

Student evaluations of instruction certainly have a role to play in the appraisal of instruction, but they should not be the sole basis for the performance evaluation. Only with corroboration from other sources of evidence (e.g., peer evaluation and cognitive testing of students) can one obtain a clear picture of an instructor's abilities in the classroom. Student evaluations form one part of the elephant. Without the others, one is not much better than blind.

Contrary to some opinions, the final word on student evaluations is not yet spoken. There is considerable need for further research concerning how students respond to questionnaires and how these questionnaires can best be used to improve instruction and to assure accountability. Attitudes and other external influences must be given further consideration by scholars.

* * *

David A. Dilts is a Professor of Labor Relations at Indiana University, Fort Wayne, Indiana and an arbitrator. He received his Ph.D. in Economics from Indiana University-Bloomington and has written extensively concerning labor relations.

Hedayeh Samavati is Assistant Professor of Economics at Indiana University-Fort Wayne. She received her Ph.D. in Economics from Iowa State University, Ames, Iowa. She has published articles concerning macroeconomics, monetary theory, statistics, and labor relations.

Mashalah Rahnama-Moghadam is Assistant Professor of Finance at Indiana University-Fort Wayne. He received his Ph.D. from Iowa State University and has published in such journals as the *Southern Economics Journal* and the *Quarterly Review of Economics and Finance.*

Lawrence J. Haber is Associate Professor of Economics at Indiana University-Fort Wayne. He received his Ph.D. in Economics from North Carolina University and has written several articles in microeconomics and labor relations.

REFERENCES

1. H. I. Mesak and L. R. Jauch, Faculty Performance Evaluation: Modeling to Improve Personnel Decisions, Decision Sciences, 22, pp. 1142-1157, 1991.
2. R. G. Schroeder, A Survey of Management Science in University Operations, *Management Science, 19*, pp. 859-906, 1973.
3. K. L. Kasten, Tenure and Merit Pay as Rewards for Research, Teaching, and Service at a Research University, *Journal of Higher Education, 55,* pp. 500-513, 1984.
4. K. O. Magnusen, Faculty Evaluation, Performance and Pay: Application and Issues, *Journal of Higher Education, 58*, pp. 516-529, 1987.
5. H. J. Brightman, Toward Teaching Excellence in the Decision Sciences, *Decision Sciences, 18*, pp. 646-661, 1987.

6. J. R. Evans, Creative Thinking and Innovative Education in the Decision Sciences, *Decision Sciences, 17,* pp. 250-262, 1986.
7. J. A. Centra, *Determining Faculty Effectiveness: Assessing Teaching, Research, and Service for Personnel Decisions and Improvement,* Jossey-Bass, San Francisco, 1980.
8. P. Seldin, *Changing Practices in Faculty Evaluation: A Critical Assessment and Recommendations for Improvement,* Jossey-Bass, San Francisco, 1984.
9. D. P. Hoyt, Interrelationships among Instructional Effectiveness, Publication Record and Monetary Reward, *Research in Higher Education, 2,* pp. 81-89, 1974.
10. J. W. Marlin and J. F. Niss, End-of-Course Evaluation as Indicators of Student Learning and Instructor Effectiveness, *Journal of Economic Education,* Spring 1980.
11. D. J. Aigner and F. D. Thum, On Student Evaluation of Teaching Ability, *Journal of Economic Education,* Winter 1986.
12. W. E. Cashin, *IDEA Paper No. 20: A Summary of Research,* Center for Faculty Evaluation and Development, Kansas State University, Manhattan, 1988.
13. V. W. Voeks and G. M. French, Are Student Ratings of Teachers Affected by Grades?, *Journal of Higher Education,* June 1960.
14. H. Zangenehzadeh, Grade Inflation: A Way Out, *Journal of Economic Education,* pp. 217-226, Fall 1988.
15. D. A. Dilts, A Statistical Interpretation of Student Evaluation Feedback, *Journal of Economic Education,* Spring 1980.
16. J. J. Seigfried and R. Fels, Research on Teaching College Economics: A Survey, *Journal of Economic Literature,* September 1979.

Direct reprint requests to:

Dr. David Dilts
Department of Economics and Finance
Indiana University at Fort Wayne
Fort Wayne, IN 46805