Teaching Evaluation: Acknowledging the New Realities in the Modern Business School Classroom

Mark A. Serva, Ph.D., Baylor University
Mark A. Fuller, Ph.D., Baylor University
Information Systems Department
Waco, Texas 76798-8005
(254) 710-4604 (office)
(254) 710-1091 (fax)
Mark_Serva@baylor.edu

ABSTRACT

New pressures are requiring business schools to re-emphasize teaching, and design performance evaluation systems that will give faculty an opportunity to improve their teaching performance, and supervisors the ability to reward them accordingly. Within a framework of multi-rater performance evaluation, seven constructs are identified as major dimensions of classroom activity, and are then used to predict student perceived performance (SPP). Results confirm that class organization and relationship with students have a large impact on SPP, but newer constructs encompassing the effective use of media and active learning techniques are also important in explaining student perceived performance.

INTRODUCTION

Regardless of the industry, the reliable and valid assessment of employee performance is critical to a company's continued well being (Kopelman, 1982; Schuster and Zingheim, 1992). Within academic disciplines, foundational motivational theories propose that workers require appropriate assessment and feedback to be productive. Equity theory (Adams, 1963), expectancy theory (Nadler and Lawler, 1977), and reinforcement theory, (Skinner, 1969) all have implications on the impact of feedback and rewards on performance.

While the private sector is seeking to evolve its performance evaluation systems to more accurately measure employee productivity, the academic world struggles with this same problem across the dimensions of research, service, and teaching. Academic rewards are frequently based on a faculty member's research record (Gomez-Mejia and Balkin, 1992), which can be easily determined by examining some combination of the number of top-tier, second-tier, and conference publications. In weighing the importance of service, however, Hutcheson (1998) notes that "Service never appeared as a convincing factor in a decision to award tenure."

If service is a secondary consideration and research is easily quantifiable and heavily weighted, then what about teaching? Certainly a majority of faculty time at most universities is spent in the classroom or preparing for work in the classroom. If teaching is often the primary mission of a university, then why do some institutions of higher learning reward research performance over teaching? Two possible answers emerge. First, these institutions may indeed value research more than teaching, in which case the performance evaluation system may be correctly aligned. Alternatively, it may be that some institutions would like to place a greater emphasis on teaching, but cannot because the systems for measuring teaching performance are poorly developed.

This latter issue raises the second question: what systems and measures should be used to evaluate faculty over these areas? Despite the fact that student evaluations of instruction (SEI) are one of the most frequently researched performance evaluation systems, they remain problematic for a number of reasons. First, most SEI are flawed. In a study by Tagomori and Bishop (1995), over two hundred SEI forms were examined for flaws. Over twenty percent of the evaluation questions examined were categorized as ambiguous (e.g., "How clear were the aims, goals, and requirements of the course") and unclear (e.g., "The total experience was very worthwhile"), with over fifty percent being categorized as subjective (e.g., "The class understood the material"). The Tagomori and Bishop (1995) study concluded that the reliability and validity of many of the instruments in use in educational institutions should be rigorously questioned.

Ironically, universities that desire more valid methods of teaching assessment should perhaps take a lesson from their own research community. One of the reasons multi-rater, or "360" (Antonioni, 1996) performance assessment has become popular in businesses is that it reduces the bias from a single source of perceptual information. Multiple measures is a basic tenet of organizational research (Judd, Smith, and Kidder, 1991). Using a multi-rater perspective, faculty teaching performance should be assessed through self-evaluation (e.g., a teaching portfolio), peer evaluation (e.g., other faculty sitting in on classes), supervisor evaluation (e.g., annual meetings with the chair and/or dean), and student evaluations (e.g., student evaluations of instruction administered each semester or student exit interviews). For these different perspectives to be valid, however, it is important that they be collected independently. The current bias from invalid teaching assessment instruments may actually be magnified in cases where student evaluations are the predominant information source for peer and supervisor evaluation. The final performance evaluation may be traced back to biases in a SEI that was never empirically tested.

While the inefficacy of existing instruments provide sufficient reason to investigate current SEI, the timing of this research also seems especially appropriate. New pressures on business schools are requiring the development of better ways to measure teaching effectiveness. The AACSB is now asking business schools to devise methods to track performance along their mission and objectives. Publications like BusinessWeek have instituted yearly rankings of business schools based on surveys administered to alumni and recruiters. These ranking systems rely heavily on the educational reputation of the institution, as opposed to measures of faculty research productivity. The development of an empirically-tested instrument directed at the assessment of business school faculty should assist colleges and universities in surmounting these new challenges.

PRIOR RESEARCH AND HYPOTHESIS

Historically, a number of dimensions have been examined in teaching instruments. Wotruba and Wright (1975) summarized twenty-one studies that investigated the qualities of effective teaching, the results of which have been used to develop hundreds of rating forms over the years. Additional work (Coffman, 1954; Hodgson, 1958; Centra, 1973) has found that the most common dimensions used include organization, structure, or clarity; teacher-student interaction or rapport; and teaching skill, communication, or lecturing ability. Other dimensions occasionally studied include evaluation of course workload/course difficulty; evaluation of grading/examinations; evaluation of impact on students (i.e., self-rated accomplishment); and global/overall effectiveness.

In 1982 Marsh identified similar dimensions, including learning/value, enthusiasm, organization, group interaction, individual rapport, breadth of coverage, exams/grades, assignments and, workload. More recent work by Grussing et al., (1994) identified course organization, teaching ability, grading and feedback, student-instructor interaction, workload and course difficulty, enthusiasm/motivation, and knowledge of subject area as the major dimensions of classroom performance. Similarly, Tang (1997) notes that organization and clarity of presentation, teacher-student interaction or rapport; communication skill, workload or course difficulty, fairness of grading and examinations, student self-rated accomplishments, and a global student rating should be included in valid teaching instruments.

A great deal of consistency is apparent across these studies relative to the teaching dimensions identified. Recent research has identified two additional factors that play a large role in the college classroom: media use (Leidner and Jarvenpaa, 1993) and active learning (Bonwell and Eison, 1991). The inclusion of these new factors was confirmed in research by Serva and Fuller (1997), revealing class organization, media use, active learning, grading fairness, relationship with students, workload, and knowledge of the material as the principal dimensions of classroom activity.

While an understanding of the dimensions of the classroom environment are crucial as a starting point in creating an effective teaching evaluation instrument, these dimensions are most useful if we examine their relative impact on overall teaching performance. The purpose of this paper is to investigate factors that impact teaching performance as perceived by students. Since students are the consumers within higher education, such a perceptual measure can give important insight into the instructor's abilities. A salesperson may be evaluated by customers, for example, and this evaluation may be used subsequently as an indicator of the salesperson's performance. While giving insights into the salesperson's abilities, this metric has limitations. High customer satisfaction scores would be misleading, for instance, if the salesperson sells items at cost to favored customers. To obtain a true metric of the salesperson's performance, multiple perspectives are warranted, including financial goals as well as customer service goals.

Similarly, while student perceived performance is an important indicator, it is not a valid surrogate of true faculty teaching performance. For that reason, this paper does not purport that the included dependent construct measures faculty performance. Instead, the construct is intended to measure student perceived performance (SPP), which may in turn be an indicator for true faculty performance. Such a study is left for future research.

We hypothesize that teaching activities can be broken down into seven constructs—class organization, media use, active learning, grading fairness, relationship with students, workload, and knowledge of the material—and that these seven constructs significantly predict student perceived performance (SPP).

RESEARCH METHOD

The survey for this research was administered during 1996 to students attending summer school classes at a private southwestern university. Business school department chairs were asked to voluntarily participate in a research study investigating a new teaching evaluation instrument. Of the six departments (accounting, economics, finance, information systems, management, marketing), all but marketing agreed to participate.* In all, 728 surveys were collected. After incomplete surveys were removed, the final sample size was 626.

Measures

The independent latent variables in this study (class organization, media use, active learning, grading fairness, relationship with students, workload, and knowledge of the material) were taken from previous research by Serva and Fuller (1997), which was based heavily on traditional constructs of teaching (Centra, 1973; Wotruba and Wright, 1975; Marsh, 1982; and Tang, 1997) while also combining new dimensions based on media use and active learning. These independent constructs were assessed using twenty-three indicators from this previous research. For the dependent construct (student perceived performance), four items were created based upon student satisfaction with the instructor and course, as well as student perceived learning.

Data Collection Procedures

To minimize data collection difficulties, existing procedures for the collection of faculty evaluations at the business school were

* Because only two different marketing instructors were teaching during the first summer session, the marketing chair was concerned that the confidentiality of the participants might be compromised. For that reason, no data were collected from the marketing department.

used. Faculty usually reserve fifteen minutes at the end of a class period toward the end of the semester and ask for a student volunteer to administer the surveys. To keep the administration of the surveys as uniform as possible, faculty were asked to read the following statement before the surveys were distributed to students:

> "In an effort to improve the procedures in which faculty are evaluated, I ask that you complete this form evaluating the quality of the instruction in this course. It is important that you complete the form honestly and completely. While the participation in the process is voluntary, your sincere cooperation will enable us to do a better job of evaluating and improving instruction here at [name of school omitted]."

The faculty were instructed to leave immediately after reading the statement to insure that the procedure remained confidential and to minimize student apprehension.

A number of procedures were instituted to ensure the completed surveys remained anonymous. Within any organization, the collection of performance information is sensitive. To ensure that the results could not be traced back to a specific faculty member, a coding scheme was devised. Each department chair was instructed to randomly pick a number to identify each faculty member within a designated numerical range (e.g., 300-399); faculty members teaching multiple classes were assigned the same numerical identifier. The faculty member was then given the surveys, and once completed, returned them to the department chairs within sealed envelopes. Graduate students then opened up the surveys and scanned the completed surveys for student references to faculty members' names, which were then blacked out with permanent marker. None of the graduate assistants were taking summer session courses. This procedure minimized the chance that any one person had access to both the data and the faculty names.

Methods Used to Compute Reliability and Validity Levels

Generally accepted procedures were used to establish the reliability and validity of the scales. For the purposes of the analysis and to minimize scaling effects, all variables were standardized to a mean of zero and a standard deviation of 1. Reliability calculations used Cronbach's alpha (Cronbach, 1951), which is equivalent to the average of all split-half reliabilities (Judd, Smith, and Kidder, 1991). Confirmatory factor analysis was used to establish convergent and discriminant validity (Bagozzi and Phillips, 1991). Hattie (1985) states that factor analysis verifies whether or not the indicators measure the same underlying construct. Factor analysis also establishes unidimensionality, which is necessary when aggregating multiple indicators into one value (Bagozzi, 1980). Regression techniques using ordinary least squares (OLS) will be used to assess the effects of the independent variables on the dependent variable.

RESULTS

Descriptive information regarding the sample is listed in Table 2. More males (53.3%) than females (43.8%) compromise the sample. As is typical of summer courses, most students took courses that are required in their field (43.3%). A smaller number took courses required for graduation, but not in their field of study (25.0%). A large percentage of the students were seniors (60.3%); the next highest percentage was juniors (18.3%). Note that while the surveys were administered largely to undergraduate students, a significant number of graduate students also participated in the research (13.6%). The largest number of students took courses offered by the management department (31.9%), with accounting (28.4%) and information systems (24.2%) close behind. Analysis of variance tests found no significant differences in SPP for gender (p=0.57), upper/lower class status (p=0.20), and department (p=0.53). Subsequent analysis, therefore, will use the pooled sample in determining results.

TABLE 2: Description Information

Gender Distribution:	Count	Percentage
Male:	388	53.3%
Female:	319	43.8
Missing:	21	2.9
Course Type:		
Required in Field	315	43.3%
Req'd, out of Field	183	25.0
Elective in Field	113	15.5
Elective out of Field	91	12.5
Missing	27	3.7
Class Status		
Senior	439	60.3%
Junior	133	18.3
Graduate	99	13.6
Sophomore	30	4.1
Freshman	5	0.7
Missing	22	3.0
Class Departments		
Management	232	31.9%
Accounting	207	28.4
Information Systems	176	24.2
Finance	91	12.5
Economics	22	3.0

Table 3 lists the Cronbach's alpha and the factor analysis results for the independent constructs. Bagozzi (1981) recommends reliability levels of at least 0.50, and the resulting levels all exceed that value. For student perceived performance, the Cronbach's alpha of 0.87 (not listed) indicates the dependent variable scale is highly reliable.

The Kaiser (1960) criterion states that factors with eigenvalues less than 1.0 should be interpreted with caution, but this recommendation is usually applied to exploratory factor analysis; that is, when the constructs have not been defined by theory a priori. Since the supporting theory calls for seven constructs, the extraction was forced to seven factors to conform to theory. Extraction was done using principal components analysis and varimax rotation was used to increase the interpretability of the resulting factors. Table 3 illustrates the rotated factor matrix for the independent variables. Factor loadings less than 0.50 are omitted.** All factors defined a priori are significantly large and

TABLE 3: Independent Variables— Validity & Reliability Calculations

	Media Use	Student Relationships	Class Organization	Active Learning	Workload	Knowledge of the Material	Grading Fairness
X1	0.82						
X2	0.83						
X3	0.85						
X4	0.85						
X5		0.66					
X6		0.73					
X7		0.83					
X8		0.80					
X9			0.74				
X10			0.58				
X11			0.69				
X12			0.78				
X13				0.84			
X14				0.86			
X15				0.70			
X16					-0.78		
X17					-0.70		
X18					-0.73		
X19						-0.72	
X20						-0.69	
X21						-0.59	
X22							-0.61
X23							-0.63
Eigenvalues	8.20	2.05	1.67	1.46	1.23	0.91	0.88
Cronbach's α	0.91	0.82	0.82	0.82	0.68	0.72	0.52

** Only one factor loading fell within the 0.40 to 0.50 magnitude range. Question Q21 loaded -0.42 on the GF (grading fairness) construct. All other loadings not listed were below 0.40 in magnitude.

load on the hypothesized factor, indicating a strong case for construct validity. No cross-loadings are present, confirming discriminant validity; that is, that the factors each represent different theoretical constructs. A separate factor analysis was performed on the indicators for the dependent variable, student perceived performance. All indicators load on one factor, indicating that the scale is unidimensional. The loadings for SPP for the four indicators are highly significant (the loadings range from 0.79 to 0.88), and the resulting eigenvalue was 2.88.

To determine the fit of the model to the data, ordinary least squares analysis was used. The stepwise procedure was followed to ensure indicators were sufficiently significant predictors of performance. The stepwise procedure will also remove an indicator if its predictive significance drops below a defined level (usually below the 95% significance level).

The stepwise procedure entered all variables into the regression in the order listed in Table 4; no variables were removed. The initial F-test for linearity indicate that the model is, indeed, linear ($F_{7,618}$ = 181.92, p<0.001). The standard error of the model is s=0.5620, and the R2 =67.3%, indicating the model explains approximately two-thirds of the overall variance. Note that because the data have been standardized, the above coefficients are scaled in terms of standard deviation. The standardization also facilitates comparisons among the predictors. All hypothesized predictors of performance are highly significant at the p<0.01 level. Correlation between the standardized residuals and the normal probability values was 0.98, indicating that the assumption of normal residuals is met.

TABLE 4: Stepwise Regression Results

Predictor	Coefficient	Standard Dev	T-value	P-value	VIF
Constant	0.02	0.02	0.84	0.404	
Media	0.16	0.03	6.03	0.000	1.4
Student Relationships	0.18	0.03	6.22	0.000	1.6
Class Organization	0.26	0.03	8.23	0.000	2.0
Active learning	0.12	0.03	4.45	0.000	1.5
Workload	0.08	0.03	3.26	0.001	1.3
Knowledge of the Material	0.08	0.03	2.77	0.006	1.8
Grading Fairness	0.22	0.03	7.5	0.000	1.7

Model Fit Statistics:
s=0.5620 R^2 = 67.3% R^2(adjusted) = 67.0% $F_{7,618}$ = 181.92

The variance inflation factor measures the degree of multicollinearity among the independent variables. While multicollinearity does not affect a model's ability to predict, independent variables that are highly correlated can result in inflation of the coefficient's standard error, resulting in misleading t-values and even incorrect signs on the coefficients (Neter, et al., 1989). Neter, et al. (1985) state that the maximum VIF in the model should be less than ten. All predictors in the model easily pass this standard, so multicollinearity does not appear to be a problem in the model.

DISCUSSION AND SUMMARY

This study examined the dimensions of teaching within a business school environment, and the strength of those dimensions in the prediction of student perceived performance. It was hypothesized that the constructs defined a priori would be significant predictors of student perceived performance. Data analysis supported the predictive ability of all the hypothesized constructs.

The strongest predictor of student perceived performance appears to be class organization. The creation of a well-defined class schedule, class objectives, and assignments may assist students in structuring their time as well as their thinking. Student may not appreciate the uncertainty of ad hoc classes; in fact, students frequently chide faculty for ambiguous assignments and not communicating clearly what is expected of them within the course. As mentioned earlier, however it should not be assumed that SPP is the same as actual performance or student learning. In some instances the creation of ambiguous situations may actually enhance learning, and thus presumably instructor performance (despite the fact the students may not immediately appreciate the value of such a learning exercise). By creating a situation where the student must not only find the answer but the question, faculty can encourage unstructured thinking within the classroom.

Relationship with students was also found to be a major predictor of SPP. A frequent complaint among faculty is the belief that students prefer to be entertained rather than taught. In such a situation, it seems likely that students would like their professor, and therefore it would be easy to dismiss the strength of a student relationship construct. The questions making up this construct, however, tell a different story. This construct emphasizes the importance of helping students and treating students with respect. Previous research (Wotruba and Wright, 1975) has found that instructors who create a comfortable environment also create an environment that is more conducive to learning.

The relative strength of the new constructs was encouraging. Media use was found to be third in importance, indicating the selection and effective use of communication tools are not only important, but may improve the students' educational experience. Leidner and Jarvenpaa (1993) found that the effective use of computer-based technology can enhance learning in the classroom. They also found, however, that how the technology was used was also important. Classes that used computers simply as a presentation device saw no advantage over overhead projectors or other traditional display media. When the technology is used as an analysis and discussion tool, however, the classes experienced more stimulating discussions that involved more complex thought processes—such as analysis—than simple processes—such as memorization. It is possible, therefore, that classes integrating new technologies (such as internet-based discussion groups, chat sessions, and electronic brainstorming) can result in stronger teaching perceptions by students.

Active learning, another new construct focused on involving students in the classroom learning experience, was also a significant predictor of SPP. The strength of the active learning construct supports a constructivist model of learning, where the classroom environment is learner-centered (O'Loughlin, 1992) as opposed to an objectivist model, where the classroom is instructor-centered (Jonassen, 1993). Recent articles highlight the importance of involving students in the learning experience, and demonstrate that the passive lecture environment is a poor method of communicating information. While potentially having a large impact, classes that cognitively engage students can be more difficult to conceive, plan, and implement. Students may realize this, and appreciate the added attention and effort that they receive during such classes.

While significant, grading and workload were less important predictors of SPP. In their meta-analysis of twenty-one research studies, Wotruba and Wright (1975) found grading fairness to be sixth on the list of most commonly cited teaching criteria, behind such factors as course organization and attitudes toward (or relationships with) students. This factor may be a "necessary, but not sufficient" condition for student perceived performance to occur. Students may feel that faculty who are not clear in their grading criteria deserve lower SPP ratings. The positive correlation for workload may need some clarification. Classes that require more work may also enable students to increase their perceived learning. Indeed, while it is clear some students avoid classes that require much work, it is also clear that students will not tolerate classes that require too little of them.

Consistent with previous research findings, the instructor's knowledge of the material is a significant—but weak—predictor of student perceived performance. While in extreme cases students may be able to detect low domain knowledge, for the most part students are not qualified to determine whether or not an instructor is knowledgeable. A previous study by Naftulin et al. (1973) exposed professional educators and students to a highly-entertaining lecture by a professional actor that contained little educational value. The study found that an instructor's knowledge of the material was not a significant predictor of instructor performance, even when assessed by the professional educators: "Given a sufficiently impressive lecture paradigm, an experienced group of educators participating in a new learning situation can feel satisfied that they have learned despite irrelevant, conflicting, and meaningless content conveyed by the lecturer" (p. 634). This construct, therefore, probably taps into the instructor's clarity of presentation and ability to enunciate and explain important class terms and concepts.

This research contributes to the existing work in faculty performance evaluation in two primary ways. First, this work focused on the development of a valid and reliable instrument, using a robust statistical methodology. All scales were found to be reliable, and factor analysis was used to confirm the construct and discriminant validity of the independent and dependent measures. The resulting regression model explained approximately two-thirds (R2=0.67) of the variance in the student perceived performance measure, a fairly high level in organizational research. The fact that discriminant validity was strong and that the new constructs of media use and active learning were significant indicates that the constructs add important information in the prediction of student perceived performance. In fact, as mentioned above, media use and active learning were more predictive of SPP than all but class organization and relationship with students. The strength of these relationships substantiates the necessity for including these factors into student evaluation of instruction instruments.

Second, this research advocates the inclusion of two new teaching performance dimensions: active learning and media use. Given the advances in educational technology, we believe that the construct dealing with the effective use of media to be of particular value, especially as we look at education in the 21st century. Business schools have started incorporating a variety of new media both to change access to educational materials and to change the educational process itself. Audio and video teleconferencing technologies, presentation software, group decision support systems, simulation and modeling software, and the Internet are all being incorporated into educational environments. Distance learning is also becoming a predominant educational delivery form. The inclusion of a teaching dimension that is consistent with use of new media is necessary for our complete understanding of the teaching environment.

Along with the strengths, the limitations of this research must be considered in interpreting the results. The data for this study were collected from one university during summer coursework. While the number of students in the sample was high, the generalizability of this sample to non-summer students, or to other universities in different areas of the country, raises external validity issues. Future research could address this deficiency through a cooperative study done at several universities in different regions of the country during either the fall or spring semester. There was also initial concern on the proportion of the sample that was graduating seniors. Would these seniors rate faculty unusually high (because they were happy to be leaving) or unusually low (because they were likely to be involved in required courses which could impact their earning potential, thus leading them to be more demanding)? Testing for differences between lower and upperclassmen, however, yielded no significant difference. While these preliminary statistical tests indicated no bias, future research should examine faculty performance across undergraduate, MBA and executive MBA programs for additional clarity on this issue.

One of the key points of this paper is that SEI are only one part of an overall teaching performance portfolio, and that we need to show caution in how we apply the knowledge gained from these instruments. As discussed above, these SEI are frequently used as the primary information source in supervisor evaluations. Any biases inherent in these instruments are thus passed along potentially infecting supervisor evaluations as well. Administrators need to be cognizant of the need for multi-rater methods (i.e. the 360 feedback approach) in assessing teaching performance. Supervisor, peer, and self evaluations should be arrived at independently, and then jointly considered as part of an overall assessment of teaching performance. Future research should focus on developing and integrating other components in a 360 feedback approach. If universities desire to improve teaching then they need to reward it. In order to reward teaching, robust measures need to be developed. This research is one step in that direction.

REFERENCES

Adams, J. (1963). Toward an understanding of inequity. *Journal of Abnormal and Social Psychology, 67*(5), 422-436.

Antonioni, D. (1996). Designing an effective 360-degree appraisal feedback process. *Organizational Dynamics, 25,* 24-39.

Bagozzi, R. (1980). *Causal Modeling in Marketing.* New York: Wiley & Sons.

Bagozzi, R. (1991). An Examination of the Validity of Two Models of Attitude. *Multivariate Behavioral Research, 16,* 323-359.

Bagozzi, R., Yi, Y., & Phillips, L. (1991). Assessing Construct Validity in Organizational Research. *Administrative Science Quarterly, 36,* 421-458.

Bonwell, C. C., & Eison, J. (1991). (Report No. ASHE-ERIC Higher Education Report 1). Washington, DC: The George Washington University School of Education and Human Development.

Centra, J. A. (1979). *Determining Faculty Effectiveness: Assessing teaching, research, and service for personnel decisions and improvement.* San Francisco: Jossey-Bass.

Coffman, W. (1954). Determining student's concepts of effective teaching from their ratings of instructors. *Journal of Educational Psychology, 45*, 277-285.

Cronbach, L. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika, 16*, 297-334.

Gomez-Mejia, L. R., & Balkin, D. B. (1992). Determinants of faculty pay: An agency theory perspective. *Academy of Management Journal, 35*(5), 921-955.

Grussing, P. G., Valuck, R. J., & Williams, R. G. (1994). Development and validation of behaviorally-anchored rating scales for student evaluation of pharmacy instruction. *American Journal of Pharmaceutical Education, 58*(Spring), 25-37.

Harvey, E. (1994). Turning Performance Appraisals Upside Down. *Human Resources Professional, 7*(2).

Hodgson, T. F. (1958). *The general and primary factors in student evaluation of teaching ability.* Seattle: University of Washington Press.

Hutcheson, P. (1998). Tenure: Traditions, Policies, and Practices. *The Review of Higher Education, 21*(3), 303-313.

Jonassen, D. (1993). Thinking technology. *Educational Technology*, 35-37.

Judd, C., Smith, E., & Kidder, L. (1991). *Sixth Edition: Research Methods in Social Relations.* Forth Worth, Texas: Holt, Rinehart, and Winston, Inc.

Kaiser, H. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement, 20*, 141-151.

Kopelman, R. (1982-1983). Improving productivity through objective feedback: A review of the evidence. *National Productivity Review*, 43-55.

Leidner, D. E., & Jarvenpaa, S. L. (1993). The Information Age Confronts Education: Case Studies on Electronic Classrooms. *Information Systems Research, 4*(1), 24-54.

Marsh, H. W. (1982). A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Education Psychology, 52*, 77-79.

Nadler, D., & Lawler, E. (1977). Motivation: A Diagnostic Approach. J.R. Hackman, E.E. Lawler, & L.W. Porter (editors), *Perspectives on Behavior in Organizations* (p. Ch. 3.). New York: McGraw-Hill.

Naftulin, D., Ware, J., & Donnelly, F. (1973). The doctor fox lecture: a paradigm of educational seduction. *Journal of Medical Education, 48*, 630-635.

Neter, J., Wasserman, W., & Kutner, M. (1985). *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs.* Homewood, Ill.: Richard D. Irwin, Inc.

Neter, J., Wasserman, W., & Kutner, M. (1989). *Applied Linear Statistical Models: Second Edition.* Chicago: Richard D. Irwin, Inc.

O'Loughlin, M. (1992). Rethinking science education: beyond Piagetian constructivism toward a socio-cultural model of teaching and learning. *Journal of Research in Science Teaching, 29*(8), 791-820.

Schuster, J. R., & Zingheim, P. K. (1992). *The New Pay: Linking Employee and Organizational Performance.* New York: Lexington Books.

Serva, M. A., & Fuller, M. A. (1998). The development of an instrument to assess the impact of media use and active learning on faculty performance. *Proceedings of the 1998 SigCPR Conference.*

Skinner, B. F. (1969). *Contingencies of Reinforcement.* New York: Appleton-Century-Crofts.

Tagomori, H. T., & Bishop, L. A. (1995). Student Evaluation of Teaching: Flaws in the Instruments. *The NEA Higher Education Journal, 11*(1), 63-78.

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*, 379-390.

Wotruba, T. R., & Wright, P. L. (1975). How to Develop a Teacher-Rating Instrument. *Journal of Higher Education, XLVI*(6), 653-663.

Information Systems & Computing
Academic Professionals

EDSIG
Serving Information Systems Educators

STATEMENT OF PEER REVIEW INTEGRITY

All papers published in the Journal of Information Systems Education have undergone rigorous peer review. This includes an initial editor screening and double-blind refereeing by three or more expert referees.

ISSN 1055-3096