Issues in Accounting Education
Vol. 13, No. 3
August 1998

Assessing Student Motivation to Participate in Teaching Evaluations: An Application of Expectancy Theory

Yining Chen and Leon B. Hoshower

ABSTRACT: The literature suggests that most universities in North America use student ratings of instruction as part of their evaluation of teaching effectiveness. Consequently, the active participation and meaningful input of students are critical factors in the success of such teaching evaluation systems. However, very few studies have looked into students' perception of the teaching evaluation system and their motivation to participate.

This study employs expectancy theory to evaluate some key factors that may motivate students to participate in the teaching evaluation process. This study found that students generally considered the improvement of teaching to be the most attractive outcome of a teaching evaluation system. Two other important uses of teaching evaluations were improving course content and format as well as making the results of evaluations available for students' decisions on course and instructor selection. The least important use of teaching evaluations, from the students' standpoint, is for the professor's tenure, promotion and salary raise decisions. Students' motivation to participate in teaching evaluations is also affected significantly by their expectation that they will be able to provide meaningful feedback. Since quality student input is an essential antecedent of meaningful student evaluations of teaching effectiveness, the results of this study should be considered thoughtfully as the evaluation system is designed, implemented and operated.

INTRODUCTION

In recent years, the practice of evaluating instruction and courses has burgeoned (Hepworth and Oviatt 1985). Wagenaar (1995) stated that well over 90 percent of schools use student ratings for assessing teaching. Evidence from many other studies also indicates that most universities in North America use student ratings of instruction as part of their evaluation of teaching effectiveness (Abrami 1989). Seldin (1985) found that systematic student ratings were "always used" as a component of faculty evaluation systems by more than 80 percent of the business schools surveyed. Calderon et al. (1994) found that close to 95 percent of accounting departments use student ratings of instruction and as many as

Yining Chen is an Assistant Professor and Leon B. Hoshower is a Professor, both at Ohio University.

We are grateful to the reviewers for their helpful comments. Please contact the authors for data availability.

18 percent rely exclusively on student ratings in evaluating faculty teaching performance.

Student evaluations of teaching effectiveness are commonly used to provide (1) formative feedback to faculty for improving teaching, course content and structure; (2) a summary measure of teaching effectiveness for promotion and tenure decisions; and (3) information to students for the selection of courses and teachers (Marsh and Roche 1993). Research on student evaluations of teaching effectiveness often examines issues like the development and validity of an evaluation instrument (Marsh 1987), the validity (Cohen 1981) and reliability (Feldman 1977) of student ratings in measuring teaching effectiveness, and the potential bias of student ratings (Abrami and Mizener 1983; Hofman and Kremer 1980; Tollefson et al. 1989). Very few studies, however, have examined students' perceptions of teaching evaluations and their motivation to participate in the evaluation. Since students' input is the root and source of student evaluation data, meaningful and active participation of students is essential. The usefulness of student evaluation data is severely undermined unless students are willing to provide quality input.

Expectancy theory has been recognized as one of the most promising conceptualizations of individual motivation (Ferris 1977). Many researchers have proposed that expectancy theory can provide an appropriate theoretical framework for research that examines a user's acceptance of and intent to use a system (DeSanctis 1983). However, empirical research on expectancy theory in the education context has been limited. This study uses expectancy theory as part of a student-based experiment to examine students' acceptance of and motivation to participate in a teaching evaluation system. The following section provides a review of prior research on teaching evaluation and a discussion of expectancy theory. Section three explains the research methodology. Results of the experiment are discussed in section four. Limitations, practical suggestions, and a summary are presented in section five.

THEORETICAL BACKGROUND AND SUPPORTING LITERATURE

Defining and measuring teaching effectiveness plays an important role in many of the decisions made in higher education. Marsh's (1987) literature review article shows the extent of research directed toward the subject of teaching effectiveness. Typically, teaching effectiveness is measured through some form of student questionnaire that has been specifically designed to measure observed teaching styles or behaviors (Wright and O'Neil 1992).

Since student ratings are used as the primary measure of teaching effectiveness, active participation and meaningful input from students are critical factors in the success of a teaching evaluation system. Several studies in the education area have observed a significant linkage between student attitudes toward the evaluation of teaching effectiveness and the success of a teaching evaluation system (Douglas and Carroll 1987; Hofman and Kremer 1980; Marsh 1984, 1987; Tom et al. 1990). However, very few studies have analyzed the factors that influence the students' attitudes toward teaching

evaluations and the relative importance of these factors. Likewise, few studies have examined the behavioral intention of students participating in the evaluation of faculty teaching effectiveness.

Student evaluations of teaching effectiveness have traditionally served two functions—as formative and summative measurements of teaching. One formative use of student evaluations is as feedback to instructors who wish to modify their teaching practices. Many studies examined the usefulness of teaching evaluations in improving teaching performance (Arubayi 1987; Divoky and Rothermel 1989; Marsh and Roche 1993; Theall and Franklin 1991; Wilson 1986). Teaching evaluations are also used to improve course content, format and structure. Studies that examined the course improvement aspect of teaching evaluations include Driscoll and Goodwin (1979) and Simpson (1995).

The summative function of teaching evaluations provides information for administrative decisions. In fact, most colleges and universities attach great importance to teaching performance in tenure, promotion and pay raise decisions (Cashin and Downey 1992; Centra 1994; Kemp and Kumar 1990; Lin et al. 1984). This summative function of teaching evaluations may also provide information for students' selection of instructors or courses (Marsh and Roche 1993). This function has been a subject of controversy and not yet widely adopted by many colleges and universities. In state supported institutions, teaching evaluations are publicly available information under the Freedom of Information Act. Student groups at some universities routinely request this data and disseminate it to the student body.

Considerable research has investigated the reliability and validity of student ratings. Reliability studies (Marlin and Gaynor 1989; Nimmer and Stone 1991; Scherr and Scherr 1990) generally address the question, "Are student ratings consistent both over time and from rater to rater?" On the other hand, validity studies (Byrne 1992; Howard et al. 1985; Tagomori and Bishop 1995) address the questions, "Do student ratings measure teaching effectiveness?" and "Are student ratings biased?" Although methodological problems have been identified, there seems to be some support for both the reliability and validity of student ratings. Overall, the literature supports the view that properly designed student ratings can be a valuable source of information for evaluating certain aspects of faculty teaching performance (Cohen 1981; Marsh 1984; Calderon et al. 1994).

While the literature supports that students can provide valuable information on teaching effectiveness, given that the evaluation is properly designed, there is a great consensus in the literature that students cannot judge all aspects of faculty performance (Cashin 1983; Seldin 1993; Centra 1993). This literature indicates that "students should not be asked to judge whether the materials used in the course are up to date or how well the instructor knows the subject matter of the course" (Seldin 1993). In both instances, the students' background and experience may not be sufficient to make an accurate assessment, thus their conclusions may be invalid. Green et al. (1998) reported that, unfortunately, 60.8 percent of student evaluations

used in accounting departments contained at least one question that require students to infer beyond their background and experience.

In summary, our study investigates the impact of the potential uses of teaching evaluations upon students' motivation to participate in the evaluation process. The four uses of teaching evaluations tested by this are (1) improving the professor's teaching, (2) influencing the professor's tenure, promotion and salary raises, (3) improving the course's content and format, and (4) making these results available for students to use in the selection of courses and teachers. Among the four, the first and third are formative and the second and fourth are summative. The second objective of this study is to examine whether an inappropriately designed teaching evaluation, which—in the perception of students—hinders students from providing valid or meaningful feedback, will affect their motivation to participate in the evaluation.

Expectancy Theory

The theory of reasoned action, as proposed by Ajzen and Fishbein (1980), is a well-researched model that has successfully predicted behavior in a variety of contexts. They propose that attitudes and other variables (i.e., an individual's normative beliefs) do not directly influence actual behavior (e.g., participation), but are fully mediated through behavior intentions, or the strength of one's intention to perform a specific behavior. This would imply that measurement of behavioral intentions (motivation) to participate is a strong and appropriate predictor (rather than only attitudes) of the success of the system.

Expectancy theory is considered one of the most promising conceptualizations of individual motivation. It was originally developed by Vroom (1964) and has served as a theoretical foundation for a large body of studies in psychology, organizational behavior and accounting (Harrell et al. 1985; Brownell and McInnes 1986; Hancock 1995; Snead and Harrell 1995; Geiger and Cooper 1996). Expectancy models are cognitive explanations of human behavior that cast a person as an active, thinking, predicting creature in his/her environment. He or she continuously evaluates the outcomes of his or her behavior and subjectively assesses the likelihood that each of his or her possible actions will lead to various outcomes. The choice of the amount of effort that he or she exerts is based on a systematic analysis of (1) the values of the rewards from these outcomes, (2) the likelihood that rewards will result from these outcomes, and (3) the likelihood of reaching these outcomes through his or her actions and efforts.

According to Vroom (1964), expectancy theory is comprised of two related models: the valence model and the force model. In our application of the theory, the valence model shows that the overall attractiveness of a teaching evaluation system to a student (V_j) is the summation of the products of the attractiveness of those outcomes associated with the system (V_k) and the probability that the system will produce those outcomes (I_{jk}). Thus:

$$V_j = \sum_{K=1}^{n} (V_k I_{jk})$$

where:

V_j = the valence, or attractiveness, of a teaching evaluation (outcome j – first-level outcome);

V_k = the valence, or attractiveness, of outcome k (second-level outcome); and

I_{jk} = the perceived probability that the teaching evaluation will lead to outcome k.

In our case, the four potential outcomes (i.e., k = 4) are the four uses of teaching evaluations that were discussed in the previous section. They are (1) improving the professor's teaching; (2) influencing the professor's tenure, promotion and salary raises; (3) improving the course's content and format; and (4) making these results available for students to use in the selection of courses and teachers.

The force model shows that a student's motivation to exert effort into a teaching evaluation system (F_i) is the summation of the products of the attractiveness of the system (V_j) and the probability that a certain level of effort will result in a successful contribution to the system (E_{ij}). Thus:

$$F_i = \sum_{J=1}^{n} (E_{ij} V_j)$$

where:

F_i = the motivational force to participate in a teaching evaluation at some level i;

E_{ij} = the expectancy that a particular level of participation (or effort) will result in a successful contribution to the evaluation; and

V_j = the valence, or attractiveness, of the teaching evaluation; derived in the previous equation of the valence model.

In summary, each student first uses the valence model and then the force model. In the valence model, each participant is given the outcomes of a teaching evaluation system (e.g., improved teaching, rewarding of effective teaching, improved course content and availability of results for students' decision making) and the subjective probability that outcomes will occur. Next, by placing his or her own intrinsic values (or weights) on the various outcomes, each student evaluates the overall attractiveness of the teaching evaluation system. Finally, the student uses the force model to determine the amount of effort he or she is willing to exert in the evaluation process. This effort level is determined by the product of the attractiveness generated by the valence model (above) and the likelihood that his or her effort will result in a successful contribution to the system. Based on this systematic analysis, the student will determine how much effort he or she would like to exert in participating in the evaluation system.

RESEARCH METHOD

Subject Selection

This study was conducted at a middle-sized (15,000 to 20,000 total enrollment), midwestern university. Five different professors, three males and two females, allowed us to administer this instrument in their nine class sections. There were four sections of managerial accounting (principles), three sections of personal income tax, and two sections of intermediate accounting. The instrument was administered at the beginning of a regularly scheduled class around the middle of the quarter to all the students who were present on that particular day. We explained the use of the instrument, read the instruction page to the students, and then asked the students to complete

the instrument. The entire process took between 15 and 20 minutes.

We divided the subjects into two groups, accounting majors and nonaccounting majors, to examine whether the results of the accounting majors would extend to students with different academic training. Believing that there would be greater internal validity in the comparison of the responses of the accounting students and the responses of the nonaccounting students if both groups had similar class rank, we eliminated the six sophomores in the sample.[1] We also eliminated four instruments which were incomplete. This resulted in 190 usable instruments composed of 92 junior and senior accounting majors and 98 juniors and seniors with majors other than accounting. Among the 98 nonaccounting majors, there were 49 from the College of Business, 24 from the College of Communications, 14 from the College of Health and Human Services, eight from the College of Arts and Sciences, and three from the College of Engineering. We also compared the responses of the accounting and nonaccounting groups with the six sophomores in the sample. The inclusion of these six students did not meaningfully alter the results.

Judgment Exercise

The within-person or individual focus of expectancy theory suggests that appropriate tests of this theory should involve comparing measurements of the same individual's motivation under different circumstances (Harrell et al. 1985; Murky and Frizzier 1986). In response to this suggestion, this study incorporates a well-established within-person methodology originally developed by Stahl and Harrell (1981) and later proven to be valid by other studies in various circumstances (e.g., Snead and Harrell 1995; Geiger and Cooper 1996). This methodology uses a judgment modeling decision exercise that provides a set of cues that an individual uses in arriving at a particular judgment or decision. Multiple sets of these cues are presented with each representing a unique combination of strengths or values associated with the cues. A separate judgment is required from the individual for each unique combination of cues presented.

We employed a one-half fractional factorial design[2] using the four second-level outcomes shown prior to Decision A at two levels (low = 10 percent and high = 90 percent). This resulted in eight different combinations of the second-level outcomes ($2^4 \times \frac{1}{2} = 8$ combinations). Each of the resulting eight combinations were then presented at two levels (10 percent and 90 percent) of expectancy to obtain 16 unique cases (8 combinations × 2 levels of expectancy = 16

[1] The normal sequence of courses for business majors is financial principles in the fall quarter of the sophomore year, and managerial principles in the winter quarter of the sophomore year. Consequently, there would be very few business sophomores taking managerial accounting during the fall quarter. During the fall quarter, these classes are populated by junior business majors and junior and senior non-business majors who are either fulfilling the accounting requirement of their major or working toward a minor in business.

[2] According to Montgomery (1984, 325), "if the experimenter can reasonably assume that certain high-order interactions are negligible, then information on main effects and low-order interactions may be obtained by running only a fraction of the complete factorial experiment." A one-half fraction of the 2^4 design can be found in Montgomery (1984, 331–334). Prior expectancy theory studies (e.g., Burton et al. 1992; Snead and Harrell 1995) also used one-half fractional factorial design.

cases). This furnished each participant with multiple cases that, in turn, provided multiple measures of each individual's behavioral intentions under varied circumstances. This is a prerequisite for the within-person application of expectancy theory (Snead and Harrell 1995).

In each of the 16 cases, the participants were asked to make two decisions. The first decision, Decision A, corresponded to the V_j in the valence model and represented the overall attractiveness of participating in the evaluation. Each participant was given the four possible outcomes of the evaluation system. Each case assigned the likelihood (10 percent or 90 percent) that each of the four second-level outcomes (V_k) would result from their participation. (The instructions and a sample case are provided in the appendix.) As mentioned earlier, the four second-level outcomes are (1) improving the professor's teaching, (2) influencing the professor's tenure, promotion and salary raise, (3) improving the course content and format, and (4) making the results available to students.

The second decision, Decision B, corresponded to F_i in the force model and reflected the strength of a participant's motivation to participate in the evaluation, using (1) the attractiveness of the evaluation (V_j) obtained from Decision A and (2) the expectancy (E_{ij}) that, if the participant exerted a great deal of effort, he or she would be successful in providing meaningful or useful input to the evaluation process.

We used 11-point response scales with ranges of –5 to 5 for Decision A and 0 to 10 for Decision B. Negative five represented "very unattractive" for Decision A and positive five represented "very attractive." For Decision B, zero represented "zero effort" and ten represented a "great deal of effort."

There is a problem with applying the expectancy theory model to teaching evaluations. Expectancy theory holds that an individual may devote a great deal of effort toward achieving an outcome but, despite this best effort, he or she may not be able to achieve the desired outcome. It may be that in the student's perception, the "successful" completion of the evaluation is trivial. All one has to do is fill in the multiple-choice grid. Likewise, the range of effort that may be devoted to completing the evaluation may not be apparent. Consequently, in the "Further Information" between Decision A and Decision B, we provided a situation in which the student was told that the hypothetical course evaluation contained "several open-ended essay questions that will require a great deal of effort for you to complete." Furthermore, we told the student that "despite your best efforts your feedback may not be helpful to the reader." This added the necessary uncertainty about the reward of effort, as well as providing a feeling that the required effort could be considerable. The students were further reminded that their participation in the evaluation is voluntary and they are free to decide to what extent they would participate in the evaluation.

It is quite common that open-ended questions are enclosed in a course evaluation to allow students to express an unconstrained opinion about some aspect of the class and/or instructor. Such questions usually provide important diagnostic information and insight for the formative evaluation of the course and

instructor (Calderon et al. 1996). Though important, open-ended questions are more difficult to summarize and report. Our instrument explained that the reader could misinterpret the evaluator's feedback on the essay. Likewise, the data from multiple-choice questions could be difficult to interpret or meaningless if the questions are designed poorly. Therefore, despite his or her efforts, the student may not be successful in contributing meaningfully to the evaluation process.

We created two different instruments; each had the order of the cases determined at random. The two instruments were distributed to every other student. We compared the average R^2s from the two random order versions and found no significant difference between them. This result implies that there is no order effect in our experimental design.

Experimental Controls

We used Pearson's correlations between R^2 values of valence and force models and selected demographic information (gender, GPA, impression toward professors, perception about the evaluation system) to test the associations between the empirical results and participants' background. The impression toward professors and perception about the evaluation system were measured by two 11-point scale demographic questions. Participating students were asked: "In general, how do you describe the professors you have had at this institution?" and "What is your general impression about the course evaluation system?"

The participants were asked to evaluate the 16 hypothetical cases (evaluation systems) presented to them instead of the evaluation systems they have experienced before. Therefore, the participants' background should not affect their responses to these individual cases. Nonsignificant correlations between participants' background (i.e., gender, GPA, impression toward professors, perception about the evaluation system) and R^2 values of valence and force models would indicate that the subjects are able to evaluate the proposed systems objectively without bias. This would support our argument that the subjects we used are appropriate for this study.

RESULTS

Valence Model

Through the use of multiple regression analysis, we sought to determine each student's perception of the attractiveness of participating in the evaluation. Decision A (V_j) served as the dependent variable, and the four second-level outcome instruments (V_k) and the probability of their occurrence (I_{jk}) served as the independent variables. The resulting standardized regression coefficients represent the relative importance (attractiveness) of each of the second-level outcomes to each participant in arriving at Decision A. The mean adjusted-R^2 of the regressions with the mean standardized betas of each outcome are presented in table 1.[3]

As indicated in table 1, the mean R^2 of the individual regression models is .7397 for the accounting group and .7377 for the nonaccounting group. There is no meaningful difference between the mean R^2 between the two groups (p-value of t-test is .9332). The

[3] Detailed regression results for each participant are not presented but are available from the authors.

TABLE 1
Valence Model Regression Results[a]

	n	Mean	Standard Deviation	Range	Frequency of Significance at .05 Level
Group I: Accounting					
Adjusted R^2	92	.7397	.1514	.2294 to .9580	88/92
Standardized Beta Weight					
V1	92	.5250	.1344	.1889 to .7939	88/92
V2	92	.3089	.2506	−.4139 to .7959	55/92
V3	92	.4126	.1763	−.1327 to .7405	71/92
V4	92	.3519	.1837	−.1539 to .7324	65/92
Group II: Non-Accounting					
Adjusted R^2	98	.7377	.1751	−.1364 to .9794	95/98
Standardized Beta Weight					
V1	98	.4641	.2034	.0166 to .9880	80/98
V2	98	.2774	.2394	−.3629 to .8154	55/98
V3	98	.3827	.1919	−.2296 to .8158	74/98
V4	98	.4243	.2567	−.1096 to .9167	78/98

[a] Results (i.e., mean, standard deviation, range, and frequency of significance at .05) of individual within-person regression models are reported in this table.

V1 = valence of teaching improvement
V2 = valence of tenure and promotion decisions
V3 = valence of course improvement
V4 = valence of result availability

mean R^2 represents the percentage of total variation in responses, which is explained by the multiple regression. Thus, these high mean R^2s indicate that the valence model of expectancy theory explains much of the variation in students' perception of the attractiveness of participating in a teaching evaluation. The absence of a difference between the groups implies that the valences of students do not differ with their major.

The standardized betas of V1, V2, V3 and V4 are significant, at the .05 level, for more than half of the individuals in both groups. This implies that all four of the secondary outcomes were important factors, to a majority of the individuals, in determining the attractiveness of a teaching evaluation system. Although all four factors were important, some factors were more important than others. It is the *mean* of these standardized betas that explains how students, on average, assess the attractiveness of potential outcomes resulting from a teaching evaluation system. Accordingly, we tested the hypothesis that $\mu_{V1}=\mu_{V2}=\mu_{V3}=\mu_{V4}$. This hypothesis was rejected at the .0001 level for both the accounting and the nonaccounting groups. Further investigation reveals that the participants, on average, placed the highest valence on the outcome V1 (teaching improvement). The strength of the other valences, in

descending order, were V3 (course improvement), V4 (availability of results) and V2 (tenure and promotion decisions) for the accounting group. The strengths of the valences of V3 and V4 reverse for the nonaccounting group, showing that this group finds V4 (the availability of results) to be more attractive than V3 (course improvement).

These results imply that both accounting and nonaccounting students believe that improving teaching (V1) is the most attractive outcome of a teaching evaluation system and that influence on the professor's tenure, promotion and salary raise (V2) is the least attractive outcome. The relative attractiveness of improving course content and format (V3) and using evaluation results for selecting courses and instructors (V4) differs between the two groups.

Panel A of table 2 summarizes the ranking of the second-level outcomes for both groups, and the statistical testing on the magnitudes of standardized betas of these outcomes. For the accounting group, the standardized betas are significantly different between V1 and V3, and V3 and V4, but not significant between V4 and V2. These results indicate that accounting students have several distinct preferences on the outcomes of the course evaluation. They consider improving teaching (V1) significantly more attractive than improving course content (V3). Likewise they consider improving course content to be significantly more attractive than having the results available for professor selection (V4). The difference between the availability of results (V4) and influence on the professor's tenure, promotion and salary raise (V2) is not significant.

In contrast, the standardized betas for V1 (improving teaching), V4 (results availability) and V3 (course improvement) of the nonaccounting students are not significantly different. This implies that nonaccounting students' preferences for these three evaluation outcomes are not as distinct as the preferences of the accounting students. Alternatively stated, the nonaccounting students considered the importance of these three outcomes to be somewhat similar, while the accounting students had more distinct hierarchical preferences. However, non-accounting students placed significantly less importance on the evaluation's impact on the professor's tenure, promotion and salary raise (V2), implying that this outcome was of less importance to them than the other three outcomes.

We also used t-tests to investigate whether there is a difference between the accounting and nonaccounting groups in their perception of the attractiveness of the four second-level outcomes. The results are reported in panel B of table 2. We found that the two groups did not view the attractiveness of V2 and V3 differently. The p-values for the t-tests were .3779 and .2670 for V2 and V3, respectively. Although both groups considered V1 (improving teaching) to be the most attractive outcome, the accounting group considered V1 significantly more attractive than the nonaccounting group. The nonaccounting group considered V4 (the results made available for students) as an outcome significantly more attractive than did the accounting group.

Panel C of table 2 presents the comparison between male and female students in perceiving the four

TABLE 2
Equality Tests

Panel A: Ranking of Second-Level Outcomes and Equality Tests on Standardized Betas

Ranking (From High to Low)	Mean of Standardized Betas	p-value of t-test
Accounting Group:		
V1	.5250	
V3	.4126	.0001 (V1 vs. V3)
V4	.3519	.0097 (V3 vs. V4)
V2	.3089	.2481 (V4 vs. V2)
Nonaccounting Group:		
V1	.4641	
V4	.4243	.3365 (V1 vs. V4)
V3	.3827	.2563 (V4 vs. V3)
V2	.2775	.0027 (V3 vs. V2)

Panel B: Equality Tests on Standardized Betas of Second-Level Outcomes Between Accounting and Nonaccounting Groups

Second-Level Outcome	Mean of Standardized Betas		p-value of t-test
	Accounting	Nonaccounting	
V1	.5250	.4641	.0166
V2	.3089	.2775	.3779
V3	.4126	.3827	.2670
V4	.3519	.4243	.0275

Panel C: Equality Tests on Standardized Betas of Second-Level Outcomes Between Male and Female Students

Second-Level Outcome	Mean of Standardized Betas		p-value of t-test
	Male	Female	
V1	.5024	.4829	.4479
V2	.3097	.2722	.2946
V3	.3598	.4424	.0020
V4	.4050	.3702	.2932

V1 = valence of teaching improvement
V2 = valence of tenure and promotion decisions
V3 = valence of course improvement
V4 = valence of result availability

second-level outcomes. The results indicate that there was no significant difference on their weight to the second-level outcomes due to gender except for V3. Female students considered improving course content as a more attractive outcome of course evaluations than did male students.

Force Model

We then used multiple regression analysis to examine the force model

(Decision B) in the experiment. The dependent variable is the individual's level of effort to participate in the evaluation (F_i). The two independent variables are (1) each student's perception about the attractiveness of the system (V_j) from Decision A, and (2) the expectancy information (E_{ij} = 10 percent or 90 percent) which is provided by the "Further Information" sentence of the test instrument (see the appendix). The force model results are summarized in table 3.

The mean R^2s (.7885 and .7989) indicate that the force model sufficiently explains the students' motivation of participating in the evaluation system. The mean standardized regression coefficient B1 indicates the impact of the overall attractiveness of the evaluation (V_j), while B2 indicates the impact of the expectation that a certain level of effort leads to successful participation in the evaluation. Our results found no significant difference between the mean standardized betas of B1 and B2 for either group of students. The p-values of these t-tests were .1792 and .8064 for the accounting and non-accounting groups, respectively. We tested the differences between the accounting and nonaccounting groups on their perception of the importance of B1 and B2 and found no significant difference at the .05 significance level. The p-values were .0920 and .6576 for B1 and B2, respectively. We also tested for differences between male and female students on their perception of the importance of B1 and B2 and found no significant difference. The p-values were .7382 and .8189 for B1 and B2, respectively. These results imply that both factors, the attractiveness of the evaluation system (B1) and the likelihood that the student's efforts will lead to success (B2), are of similar importance to the student's motivation.

TABLE 3
Force Model Regression Results[a]

	n	Mean	Standard Deviation	Range	Frequency of Significance at .05 Level
Group I: Accounting					
Adjusted R^2	92	.7885	.1587	.1502 to .9813	91/92
Standardized Beta Weight					
B1	92	.6103	.2570	−.8350 to .9889	87/92
B2	92	.5444	.2738	−.2531 to .9677	76/92
Group II: Nonaccounting					
Adjusted R^2	98	.7989	.1604	.2102 to 1.000	94/98
Standardized Beta Weight					
B1	98	.5435	.3019	−.1264 to 1.000	76/98
B2	98	.5588	.3346	−.1679 to 1.000	77/98

[a] Results (i.e., mean, standard deviation, range, and frequency of significance at .05) of individual within-person regression models are reported in this table.
B1 = weight placed on attractiveness of the evaluation
B2 = weight placed on the expectancy of successfully participating in the evaluation

Experimental Controls

Table 4 presents Pearson's correlations between R^2 values of valence and force models and selected demographic information (i.e., gender, GPA, impression toward professors, perception about the evaluation system). As shown in the two right-hand columns, there was no significant correlation (at the .05 significance level) for either the accounting or nonaccounting group between the participants' R^2 values and their gender, GPA, impression toward professors, or their perception about the evaluation system. These results suggest that neither the students' perception of the attractiveness of the evaluation system nor their motivation to participate is correlated with their background or with their prior experience relative to the evaluation system. These results also support our argument that the subjects we used are appropriate for this study because neither their background nor their prior experience with professors and teaching evaluations affected their perceptions of the evaluation systems tested in the questionnaire.[4]

An interesting finding is that for the accounting group the correlation between the students' overall impression of professors and the students' impression about the course evaluation systems is positive (.4074) and significant ($p < .0001$). This implies that accounting students who have more positive comments about the professors that they have had in the university generally consider the course evaluation system more useful. However, this result may be sample-specific, as it does not hold for the nonaccounting group.

LIMITATIONS, SUGGESTIONS AND SUMMARY

Limitations and Future Research

Some limitations of this study need to be discussed. First, the selection of subjects was not random. Students became subjects by virtue of being present on the day that their class was surveyed. The selection of classes was arbitrary. Consequently, caution should be used in generalizing the results to other groups and settings without further research. Second, an experimental task was used in this study and the subjects' responses were gathered in a controlled environment rather than in a real-world setting, although sitting in a classroom completing a teaching evaluation and sitting in a classroom completing an instrument about teaching evaluations are similar activities. Third, students were not given the opportunity for input on the outcomes that motivate them to participate in a teaching evaluation. In the instrument, four possible outcomes were given to the students. It is possible that other possible outcomes of teaching evaluations may have a stronger impact on students' motivation than the four outcomes used in this study. Future research can solicit input from accounting students on what, specifically, they see

[4] It is reasonable to expect an association between someone's prior experience with an evaluation system and his or her motivation to participate in that particular system. However, the participants were asked to evaluate the 16 proposed cases (evaluation systems) but not the system they have experienced. Therefore, the nonsignificant correlations indicate that the subjects were able to evaluate the proposed systems objectively without bias, thus supporting our argument that the subjects we used are appropriate for this study.

TABLE 4
Pearson's Correlation Coefficients/p-Values

	GPA	Impression Toward Professors	Usefulness of Evaluations	Adj-R^2 Force	Adj-R^2 Valence
Group I: Accounting					
Gender	−0.0992	−0.1050	−0.1144	0.0385	−0.0072
	(0.3494)	(0.3193)	(0.2775)	(0.7154)	(0.9459)
GPA	1.0000	0.1904	0.1718	0.0470	0.0615
	(0.0000)	(0.0707)	(0.1035)	(0.6584)	(0.5625)
Impression Toward Professors		1.0000	0.4074	0.1236	−0.1234
		(0.0000)	(0.0001)	(0.2403)	(0.2411)
Usefulness of Evaluation			1.0000	−0.0696	−0.0221
			(0.0000)	(0.5099)	(0.8346)
Group II: Nonaccounting					
Gender	−0.1414	0.0295	−0.0382	0.0392	−0.0291
	(0.1671)	(0.7735)	(0.7100)	(0.7018)	(0.7763)
GPA	1.0000	0.1601	−0.0175	0.0122	−0.1254
	(0.0000)	(0.1173)	(0.8653)	(0.9058)	(0.2212)
Impression Toward Professors		1.0000	0.0428	0.0729	−0.1320
		(0.0000)	(0.6770)	(0.4756)	(0.1950)
Usefulness of Evaluation			1.0000	0.0122	−0.1254
			(0.0000)	(0.9058)	(0.2212)

or would like to see as the outcomes of an evaluation system. Fourth, the extreme levels of instrumentality and expectancy (10 percent and 90 percent) were used in the cases. This did not allow us to test for the full range within the extremes. In another sense, such extremes may not exist in actual practice. Fifth, all subjects came from only one institution, which may limit the applicability of the results to other academic environments. Extensions can be made by future studies to examine the effect of academic environments on the results of this study.

Practical Suggestions

Toward the goal of motivating students to participate in the teaching evaluation process, we make the following practical suggestions. First, consider listing prominently the uses of the teaching evaluation on the evaluation instrument. This will inform the students of the uses of the evaluation. If these uses are consistent with the uses that students prefer (and they believe that the evaluations will truly be used for these purposes), the students will assign a high valence to the evaluation system. The next step is to show students that their feedback is really used. Accomplishing this will increase their subjective probabilities of the secondary outcomes that are stated on the evaluation. It would also increase their subjective probabilities that they will be successful in providing

meaningful feedback, since they will see that previous feedback has been used successfully. Thus, their force or motivation to participate will be high.

One way of showing students that their feedback has been used successfully is to require every instructor to cite on the course syllabus one recent example of how student evaluations have helped improve this particular course or have helped the instructor to improve his or her teaching. This seems like a low cost, but highly visible way to show students the benefits of teaching evaluations. It may also have the salutary effect of encouraging faculty to ponder the information contained in student evaluations and to act upon it. This research shows that students perceive the most attractive outcome of an evaluation system as improved teaching, while improvement of the course is ranked second by accounting majors. Thus, students who believe that their feedback on evaluations will improve teaching or the course, or both, should be highly motivated to provide such feedback.

Another way of showing students that their input on teaching evaluations is being used is to disseminate the information to students. This can be done through the student senate, the student newspaper, or through the university's web site.

Summary

The expectancy model used in this study provides a good overall explanation of a student's motivation to participate in the evaluation of teaching effectiveness. The valence model significantly explains a student's assessment of the attractiveness of a teaching evaluation system. Further, the force model provides a good explanation of a student's motivation to participate in the teaching evaluation. By the successful application of expectancy theory, this study provides a better understanding of the behavioral intention (motivation) of students' participation in the teaching evaluation process.

Our empirical results show that students have strong preferences for the uses of teaching evaluations and these preferences are remarkably consistent across individuals. Since quality student participation is an essential antecedent of the success of student evaluations of teaching effectiveness, this knowledge of student motivation (V1 through V4) must be considered thoughtfully when the system is implemented. If, however, students are kept ignorant of the use of teaching evaluations or if teaching evaluations are used for purposes which students do not value or if they see no visible results from their participatory efforts, they will cease to give meaningful input.

APPENDIX
INSTRUCTIONS

At the end of each quarter you are asked to evaluate the courses you have taken and the professors who have conducted those courses. These evaluations may be used in various ways, such as: improving teaching; rewarding (or punishing) the professors' performance with tenure, promotion or salary increases; improving course content; and providing information to future students who are contemplating taking this course or this professor.

This exercise presents 16 situations. Each situation is different with respect to how the evaluation is likely to be used. We want to know how attractive participation in the evaluation is to you in each given situation.

You are asked to make two decisions. You must first decide how *attractive* it would be for you to participate in the evaluation (DECISION A). Next you must decide how much *effort* to exert in completing the evaluation (DECISION B). Use the information provided in each situation to reach your decisions. There are no "right" or "wrong" responses, so express your opinions freely. A sample situation is provided below. The 16 different situations start on the next page.

EXAMPLE QUESTIONNAIRE

The likelihood that your feedback:

will be taken seriously by the professor for improving his/her teaching in the future is ..... HIGH (90%)
will be taken into consideration for the professor's tenure, promotion and salary raise is ..... HIGH (90%)
will be used to improve the same course (i.e., content, format, etc.) in the future is ..... HIGH (90%)
will be provided to future students who are contemplating taking this course and this professor is ..... LOW (10%)

DECISION A: With the above outcomes and associated likelihood levels in mind, indicate the attractiveness to you of participating in the evaluation.

–5	–4	–3	–2	–1	0	+1	+2	+3	+4	+5
Very Unattractive										Very Attractive

FURTHER INFORMATION: The course evaluation contains several open-ended essay questions, which will require a great deal of effort for you to complete. (As you know, your participation in course evaluations is voluntary. Thus, you can choose to exert much effort in the hopes of providing meaningful feedback or, at the other extreme, you can do nothing.)

If you exert a great deal of effort, the likelihood that the reader will find your feedback helpful is ..... LOW (10%)[5]

DECISION B: Keeping in mind your attractiveness decision (DECISION A) and the FURTHER INFORMATION, indicate the level of effort you would exert to participate in the evaluation.

0	1	2	3	4	5	6	7	8	9	10
Zero Effort										Great Deal of Effort

[5] Despite your best efforts to articulate your feelings, the reader may misinterpret your feedback. Even the responses to multiple choice questions are difficult to interpret when the questions are designed poorly.

REFERENCES

Abrami, P. C. 1989. How should we use student ratings to evaluate teaching? *Research in Higher Education* 30 (April): 221–227.

———, and D. A. Mizener. 1983. Does the attitude similarity of college professors and their students produce "bias" in the course evaluations? *American Educational Research Journal* 20: 123–136.

Ajzen, I., and M. Fishbein. 1980. *Understanding Attitudes and Predicting Social Behavior.* Englewood Cliffs, NJ: Prentice Hall.

Arubayi, E. A. 1987. Improvement of instruction and teacher effectiveness: Are student ratings reliable and valid? *Higher Education* 16 (Fall): 267–278.

Brownell, P., and M. McInnes. 1986. Budgetary participation, motivation, and managerial performance. *The Accounting Review* 61 (October): 587–600.

Burton, F. G., Y. Chen, V. Grover, and K. A. Stewart. 1992. An application of expectancy theory for assessing user motivation to utilize an expert system. *Journal of Management Information Systems* (Winter): 183–198.

Byrne, C. J. 1992. Validity studies of teacher rating instruments: Design and interpretation. *Research in Education* 48 (November): 42–54.

Calderon, T. G., B. P. Green, and B. P. Reider. 1994. Extent of use of multiple information sources in assessing accounting faculty teaching performance. In *1994 Ohio Regional Meeting Proceedings.* American Accounting Association, Ohio Regional Meeting.

———, A. L. Gabbin, and B. P. Green. 1996. *A Framework for Encouraging Effective Teaching.* American Accounting Association, Harrisonburg, VA: James Madison University.

Cashin, W. E. 1983. Concerns about using student ratings in community college. In *Evaluating Faculty and Staff: New Directions for Community Colleges,* edited by A. Smith. San Francisco, CA: Jossey-Bass.

———, and R. G. Downey. 1992. Using global student rating items for summative evaluation. *Journal of Educational Psychology* (December): 563–572.

Centra, J. A. 1993. *Reflective Faculty Evaluation.* San Francisco, CA: Jossey-Bass.

———. 1994. The use of the teaching portfolio and student evaluations for summative evaluation. *Journal of Higher Education* 65 (September/October): 555–570.

Cohen, P. A. 1981. Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research* 51: 281–309.

DeSanctis, G. 1983. Expectancy theory as explanation of voluntary use of a decision support system. *Psychological Reports* (January): 247–260.

Divoky, J. J., and M. A. Rothermel. 1989. Improving teaching using systematic differences in student course ratings. *Journal of Education for Business* 65 (December): 116–119.

Douglas, P. D., and S. R. Carroll. 1987. Faculty evaluations: Are college students influenced by differential purposes? *College Student Journal* 21 (Winter): 360–365.

Driscoll, L. A., and W. L. Goodwin. 1979. The effects of varying information about use and disposition of results on university students' evaluations of faculty courses. *American Educational Research Journal* 16: 25–37.

Feldman, K. A. 1977. Consistency and variability among college students in their ratings among courses: A review and analysis. *Research in Higher Education* 6: 223–274.

Ferris, K. R. 1977. A test of the expectancy theory as motivation in an accounting environment. *The Accounting Review* 7: 605–614.

Geiger, M. A., and E. A. Cooper. 1996. Using expectancy theory to assess student motivation. *Issues in Accounting Education* 11: 113–129.

Green, B. P., T. G. Caleron, and B. P. Reider. 1998. A content analysis of teaching evaluation instruments used in accounting departments. *Issues in Accounting Education* 13: 15–30.

Hancock, D. R. 1995. What teachers may do to influence student motivation: An application of expectancy theory. *The Journal of General Education* (Fall): 171–179.

Harrell, A. M., C. Caldwell, and E. Doty. 1985. Within-person expectancy theory predictions of accounting students' motivation to achieve academic success. *The Accounting Review* 60 (October): 724–735.

Hepworth, D., and B. E. Oviatt. 1985. Using student course evaluations: Findings, issues, and recommendations. *Journal of Social Work Education* 21 (Fall): 105–112.

Hofman, F. E., and L. Kremer. 1980. Attitudes toward higher education and course evaluation. *Journal of Educational Psychology* 72: 610–617.

Howard, G. S., C. G. Conway, and S. E. Maxwell. 1985. Construct validity of measures of college teaching effectiveness. *Journal of Educational Psychology* 77 (April): 187–196.

Kemp, B. W., and G. S. Kumar. 1990. Student evaluations: Are we using them correctly? *Journal of Education for Business* 66 (November/December): 106–111.

Lin, Y. G., W. J. McKeachie, and D. G. Tucker. 1984. The use of student ratings in promotion decisions. *Journal of Higher Education* (September/October): 583–589.

Marlin, J. E., Jr., and P. Gaynor. 1989. Do anticipated grades affect student evaluations? A discriminant analysis approach. *College Student Journal* 23 (Summer): 184–192.

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

———. 1987. Students' evaluations of university teaching: Research findings methodological issues and directions for future research. *International Journal of Educational Research* 11: 353–388.

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

Montgomery, D. C. 1984. *Design and Analysis of Experiments.* New York, NY: John Wiley & Sons.

Murky, D., and K. B. Frizzier. 1986. A within-subjects test of expectancy theory in a public accounting environment. *Journal of Accounting Research* (Autumn): 400–404.

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

Scherr, F., and S. S. Scherr. 1990. Bias in student evaluations of teacher effectiveness. *Journal of Education for Business* 65 (May): 356–358.

Seldin, P. 1985. *Current Practices in Evaluating Business School Faculty.* Pleasantville, NY: Center for Applied Research, Lubin School of Business Administration, Pace University.

———. 1993. The use and abuse of student ratings of professors. *The Chronicle of Higher Education* (July): A40.

Simpson, R. D. 1995. Uses and misuses of student evaluations of teaching effectiveness. *Innovative Higher Education* (Fall): 3–5.

Snead, K. C., and A. M. Harrell. 1995. An application of expectancy theory to explain a manager's intention to use a decision support system. *Decision Sciences* 25 (Fall): 499–513.

Stahl, M. J., and A. M. Harrell. 1981. Modeling effort decisions with behavioral decision theory: Toward an individual differences model of expectancy theory. *Organizational Behavior and Human Performance* 27: 303–325.

Tagomori, H. T., and L. A. Bishop. 1995. Student evaluation of teaching: Flaws in the instruments. *Thought & Actions: The NEA Higher Education Journal* 11: 63.

Theall, M., and J. Franklin. 1991. Using student ratings for teaching improvement. *New Directions for Teaching and Learning* 48 (Winter): 83–96.

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

Tom, G., S. Swanson, and C. Abbott. 1990. The effect of student perception of instructor evaluations on faculty evaluation scores. *College Student Journal* 24 (September): 268–273.

Vroom, V. C. 1964. *Work and Motivation.* New York, NY: John Wiley & Sons.

Wagenaar, T. C. 1995. Student evaluation of teaching: Some cautions and suggestions. *Teaching Sociology* 64 (January): 64–68.

Wilson, R. C. 1986. Improving faculty teaching: Effective use of student evaluations and consultants. *Journal of Higher Education* 57 (March/April): 196–211.

Wright, W. A., and M. C. O'Neil. 1992. Improving summative student ratings of instruction practice. *Journal of Staff, Program, and Organizational Development* 10: 75–85.