



Are Students Telling Us the Truth? A Critical Look at the Student Evaluation of Teaching

Dennis E. Clayson & Debra A. Haley

To cite this article: Dennis E. Clayson & Debra A. Haley (2011) Are Students Telling Us the Truth? A Critical Look at the Student Evaluation of Teaching, Marketing Education Review, 21:2, 101-112, DOI: [10.2753/MER1052-8008210201](https://doi.org/10.2753/MER1052-8008210201)

To link to this article: <https://doi.org/10.2753/MER1052-8008210201>



Published online: 08 Dec 2014.



Submit your article to this journal [↗](#)



Article views: 435



View related articles [↗](#)



Citing articles: 13 View citing articles [↗](#)

ARE STUDENTS TELLING US THE TRUTH? A CRITICAL LOOK AT THE STUDENT EVALUATION OF TEACHING

Dennis E. Clayson and Debra A. Haley

Over 99 percent of business schools use student evaluation of instruction to measure teaching and classroom performance. The resultant measures are utilized in judgments of merit pay, tenure, and promotion. In such an environment, an inspection of exceptions to their assumed validity is justified. This paper investigates one such issue that is rarely reported. Simply put: to what extent are students telling us the truth when they evaluate instruction? A literature review indicates that students (1) ignore or falsify answers in light of variables considered more important, (2) give subjective impressions in response to objective questions, and (3) at times give purposefully misleading and false responses. A survey of students found that a majority knew of respondents who had falsified evaluations and that an estimated 30 percent of evaluations contain answers the students knew were not true. In light of these findings, the validity of student evaluation of teaching to improve individual instructor performance, modify curriculum, and create comparative scales to evaluate faculty is called into question.

It is generally assumed that students make honest responses when they evaluate instruction. This assumption is so prevalent that it is seldom questioned. Summaries of student evaluation of teaching (SET) research generally do not include any discussion of possible student dishonesty on any aspect of the SET process. As an example, Cashin (1995) produced a short article outlining the research on SET, which was designed for college-level instructors. He reviews literature showing that SET instruments are generally reliable and then summarizes problems with administration that may bias the results, but no discussion is held of any problems with student bias or truthfulness. While scales have been developed to measure the students' perception of instructor credibility and honesty (see McCroskey and Young 1981 for the teacher credibility scale), the extant literature simply assumes that students will be honest in their evaluations. However, there is some isolated evidence that students are not always truthful in their responses. The current paper investigates this possibility and endeavors to shed some light on this problem.

Dennis E. Clayson (Ph.D., Brigham Young University), Department of Marketing, University of Northern Iowa, Cedar Falls, IA, dennis.clayson@uni.edu.

Debra A. Haley (Ph.D., Oklahoma State University), Southeastern Oklahoma University, Durant, OK, dhaley@se.edu.

HISTORICAL INTRODUCTION

Few issues in academia have been as well researched and debated as the student evaluation of teaching. Since the 1920s (Remmers and Brandenburg 1927), there has been a constant stream of articles about the evaluations. During this period, the adoption of the resultant instruments has become almost universal. By the 1990s, four of five U.S. campuses used some form of student evaluations (Seldin 1993). In business schools, almost 100 percent now use an SET procedure to evaluate their faculty (Comm and Manthaisel 1998). On many campuses, these evaluations are the most important and, in many cases, the only measure of teaching ability that is utilized (Wilson 1998). A California dean was reported as saying, "If I trust one source of data on teaching performance, I trust the students" (Seldin 1999, p. 15).

Because the evaluations are generally included in the tenure, promotion, and merit pay decision process, it is understandable that they have been widely debated and researched. It is also understandable why issues of validity have been raised since the inception of their general utilization. In this wide debate, one area has been generally ignored. Do the instruments measure some objective reality? The lack of interest in this topic probably results from two characteristics of the evaluation process. First, although there are questions of fact on most instruments, the evaluations are generally designed to solicit students' opinions and perceptions. Some have suggested that even the answers to objective questions should be seen as

subjective impressions. As Marks concluded after his study of business students, “No matter how reliable the measures, student evaluations are no more than perceptions and impressions” (2000, p. 117). Second, some have adopted a student-oriented perspective summarized by Machina (1987), who dismissed any argument of what is being measured and seemingly took a purely consumer orientation. He stated that “when students underrate a faculty member, that means there is a breakdown in the teaching process somewhere and it does not mean that the evaluation is an inaccurate measure of the thing it is really geared to measure” (Machina 1987, p. 22).

Early work in psychology and psychiatry, however, showed that external validity was compromised when respondents made invalid observations (Orne and Holland 1968). While perceptions may be important in the SET process, one of the primary functions of the evaluation is to guide improvements in instruction. As Popham observed, “Dishonest responses could lead the teacher to make invalid inferences, which would lead to poor instructional decisions” (2001, p. 122). Invalid responses can also compromise the general level of instruction when no corrections are made or when decision makers are led to incorrect conclusions from inaccurate data. Engelland noted, “SET instruments are only effective if they assist professors in improving teaching performance by providing diagnostic information that can result in actionable changes” (2004, p. 45).

PURPOSE

The current study looks at the validity issue by first reviewing the extant literature and then following up with a survey to gauge student perceptions of the issue, all designed to investigate two specific questions:

RQ1: Are the students' perceptions valid in areas where they can be compared to subjective material?

RQ2: Do students tell the truth when they fill out the evaluations?

To the extent that the answers to these questions can be “no” raises some fundamental issues of how valid the SET process is for faculty improvement, instructional correction, and the appropriateness of applying SET when utilized in cases involving tenure, promotion, and merit pay, especially when results are used comparatively between faculty. This issue is of special interest to marketing faculty not only on a personal level but also in relation to marketing fundamentals. Within the SET process, marketing faculty become the object of what they teach, that is, customer satisfaction,

the measurement of that satisfaction, how service can be improved, and the ethics involved in the process.

REVIEW OF THE LITERATURE

The Halo Effect

The “halo effect” has been of interest in psychological and marketing research for the past half-century. It was defined by Aiken as “[r]ating an examinee high on one trait or characteristic simply because he or she rates high on other characteristics” (1979, p. 307). A strong halo effect has been found in the evaluations (Orsini 1988; Tang and Tang 1987). This helps to explain an apparent inconsistency oddity found in the validity of the instruments. Researchers have argued that while the convergent validity of SET is rather high, the instruments lack discriminate validity (Clayson 2008; Marks 2000). In other words, the outcome of the evaluations is related to other variables that would logically be assumed to be associated with the concept of “good” teaching, but the evaluations are also related to many other variables that would logically be unrelated to the concept. Much of this could be explained by the spurious associations created by halo effects. Students will associate variables that would normally be seen as randomly related because both variables pertain to an instructor whom the students associate with positive or negative attributes. As a demonstration of this, researchers looking at almost 7,000 professors found a correlation of 0.64 between the instructor’s “hotness” and the quality of the class (Felton et al. 2004). One of the authors of the current paper once found a correlation greater than 0.60 from over 500 evaluations between instructors’ handwriting and adequate office hours.

The halo effect suggests that students believe that some variables are so important that they justify ignoring or falsifying other variables. One such consideration is the student’s expected grade, which has been found to be highly related to the halo effect (Clayson 1989). Another is the students’ desire to send a message. Dunegan and Hrivnak (2003) reported that when students found an instructor whom they did not like, they carefully completed the evaluation so that their perception of overall performance and SET were correlated, but if the instructor aligned with the students’ ideal, the association between performance and the SET was not significant. In other words, under this condition, the students simply and “mindlessly” (the authors’ term) marked the form.

Students also project their own feelings and bias onto the instructor. This effect was found by Tang and Tang

(1987), who reported that students would attribute their own behavior to the instructor. Students who reported reading the textbook before coming to class rated the instructor's knowledge higher, and students associated their own level of participation and involvement in the classroom with faculty performance. Tang and Tang concluded that the evaluations might give a better indication of the student's self-concept than the instructor's actual performance. Grimes, Millea, and Woodruff (2004) found that students with an external locus of control gave their instructors lower evaluations than students with an internal orientation. Attribution effects are also strong. Benz and Blatt (1996) found in an analysis of student comments that students "satisfied" with their grades described their experiences with "I," for example, "I get good grades for doing good work." Students "dissatisfied" with their grades tended to use "he" or "she," for example, "He never gives good grades."

If the halo effect allows personal bias to be projected onto an instructor, there could be some potential civil rights and diversity issues (Clayson and Haley 2005; Haskell 1997). Studies of political science students have found that instructors who were perceived as politically similar to students rated higher. As the perceived political difference grew larger, the evaluations became progressively lower (Kelly-Woessner and Woessner 2006). Kelly-Woessner and Woessner note that at colleges with a liberal bent, "conservative, right-leaning college professors may be at a slight disadvantage" (2006, p. 500). There are a consistent 20 percent of instructors who are rated by students in the same class as the *very* best or the *very* worst teachers that a student has *ever* had (Follman 1984). One explanation of this extreme reaction may be the halo effect created by differences that could be unrelated to instruction. Instructors who are different in some way, perhaps in personality (see Clayson and Sheffet 2006), but in other instances, even in geographical or racial backgrounds, may receive lower evaluations.

Perception of Objective Measures

A number of SET instruments ask students objective questions that are then utilized as part of the judgment made of instruction. Many procedures ask students about their grades, for example, because grades have been found to have an associational relationship to the evaluations (for extensive reviews, see Johnson 2003; Marsh and Roche 2000), and almost all SETs ask the students to indicate satisfaction with learning in the class. But can these expectations be used as objective points of reference?

Some commentators have asserted that students are the best judge of what they are learning (Cruse 1987; Machina 1987), but this claim has not been widely supported by objective measures. Gremler and McCollough (2002) reported that the students' rating of their adequacy as a student, along with their satisfaction with their own effort and participation in class, were only weakly related to course evaluations. In a well-conceived and carefully executed study of over 20,000 students taking economics classes at The Ohio State University, Weinberg, Fleisher, and Hashimoto found that grades were consistently related to the evaluations, but the evaluations given in prerequisite classes were unrelated to future performance: "Once current grades are controlled, learning, as measured by future grades, is never statistically significantly related to evaluations" (2007, p. 16). In addition, none of the individual items on the evaluation that were associated with the evaluations were significantly related to learning as measured by future performance. Clayson (2009), in a meta-analysis of the literature, found no relationship between learning and SET in business classes. Johnson (2003) actually reported a negative relationship looking at a large institutional sample at Duke University, a finding reinforced by a recent report showing that students in calculus classes, where performance can be objectively measured, gave lower evaluations to instructors who produced students who performed better in subsequent classes (Carrell and West 2010).

It is known that students make consistent errors in estimating their grades, which has been interpreted as a "metacognitive" effect. Poorer students don't know what they don't know and consequently overestimate their knowledge of tested material, while better students know what they don't know and underestimate their knowledge (Grimes 2002; Kennedy, Lawton, and Plumlee 2002; Moreland, Miller, and Laucka 1981). A more recent finding indicates that this interpretation may need to be modified. Students appear to know what they don't know, but use this only as a data point, which is combined with group norms to estimate their own learning performance (Clayson 2005a).

Purposeful Misreporting

Students seem willing to report information on the evaluations that they know is not true. Whether this is malicious, or even done consciously, is not always evident. For example, reports have shown that students will evaluate class presentations and instructors that do not exist. Some evidence suggests that a majority of students will evaluate anything they are asked to judge, whether or not they

have any basis for the judgment. Reynolds (1977) reported an incident in which nearly 1,000 students completed an evaluation of a course in which there were 10 invited speakers. One speaker never appeared. Nevertheless, 80 percent of the students evaluated the nonexistent lecture, ranking it worse than six but better than three. Some of the same students were shown a film in class while others did not see the film. About 55 percent of the students who did not see the film evaluated it anyway, giving it a slightly above-average evaluation. Reynolds notes, tongue-in-cheek, an interesting strategy that could be adopted if the evaluations are to be taken at face value:

further reflection reveals a subtle means of cost cutting without curtailing academic programs. Without prior announcement, a few lectures or films can be quietly eliminated. Gradually, on a random basis, further lectures can simply disappear, first in one department and then in another. Who knows how long this may go on before anyone notices? (1977, p. 83)

There is some evidence of more purposeful dishonesty. Stanfel (1995) conducted an experiment in which responses on the evaluation were matched with objective evidence. The evaluation used by his university contained several questions that could be corroborated: "The instructor explains clearly to students how they are evaluated," and "Tests and written assignments are graded and returned in a reasonable period of time." Stanfel explained his evaluation procedures and then gave his students a quiz to test their knowledge of his instruction. All the students were able to correctly outline exactly what his evaluation procedures were. In other words, there was a documented understanding, if not a consensus, of the evaluation process. All the assignments were returned at the first possible opportunity, and the students signed documents indicating that they had received their graded work. How did the students later respond on the SET? Only 3 percent of the students "strongly agreed" that the instructor had explained his evaluation procedure; 64 percent either "disagreed" or "strongly disagreed." Only 3 percent "strongly agreed" that assignments were handed back in a reasonable period of time, even though the students had previously signed documents indicating that they had received their work at the first opportunity; over 46 percent "disagreed" or "strongly disagreed." Stanfel (1995) concluded that the students either forgot what had been clearly presented to them, that they did not understand what the evaluation was asking, or that they intentionally made false responses. He maintained that the latter is the most logical explanation. Even if Stanfel's last hypothesis was incorrect, the previous

hypotheses would call into question the accuracy of his students' responses.

Sproule (2000) recounts how 50 percent of his students in one class would not acknowledge that work was returned "reasonably promptly" even when *all* work was marked and returned during the next class period. The examples are not always negative for the instructor. An anecdotal account is warranted here. One of the authors of the current paper once taught at a small elite private school. During one term, an elective course was taught to seniors. All the students were thoroughly familiar with the instructor after over three years of interaction. Because of distance and some unusual circumstances, the instructor realized that he was showing up late to almost every class meeting. On the college's SET was an unambivalent question about meeting the class on time. As a test of the accuracy of student responses, the instructor continued to come to class a few minutes late for the remainder of the term. On the evaluation, 100 percent of the students reported that the instructor showed up to class "on time."

At the negative extreme, students have been known to threaten revenge by utilizing the SET. Dershowitz (1994) recalls a disagreement he had with a group of students; a few days later, one of the students told him that he should expect to be "savaged" in the next evaluations. One of the authors of the current paper can attest to an almost identical experience.

A survey (Clayson 2005b) of students from three universities across the United States found that 30 percent admitted to purposely inflating evaluations beyond what was deserved because an instructor gave good grades; another 30 percent indicated that they had purposely lowered evaluations below that which was deserved because the tests in the course were "too hard." Fifty percent had done one or the other. Eighty percent of the students surveyed indicated that they had knowingly given an instructor an *undeserved* evaluation for some reason.

While student dishonesty has not been discussed much in the literature, the actual SET procedure puts a high premium on discouraging dishonest behavior, ironically, however, only on the part of the instructor. Gray and Bermann (2003) review how the evaluations are administered and how the faculty is explicitly forbidden to touch the evaluation sheets after they are completed. They claim,

This procedure tells the students that the teacher is more than likely to be a cheat and a sneak, who will cook the books if given a chance. Both the students and teacher pretend not to notice the shaming involved, but it is palpable in such a situation. (Gray and Bermann 2003, p. 46)

Stanfel makes an interesting observation while comparing most institutions' policies toward test taking and the evaluations:

It would be highly inconsistent of university officials to act on the one hand as if students were likely to seize every opportunity at dishonesty and on the other, to attribute to their opinions about course and instructor quality an unimpeachable purity. (1995, p. 121)

Literature Summary

Four summary conclusions can be drawn from the review of the literature:

1. There is a strong halo effect with SET indicating that students will ignore the actual content of the questions or statements by answering them in a manner consistent with a more global student concern or issue.
2. Many items that the evaluations assume are objective facts, such as the class meeting on time, are reported as subjective perceptions by the students.
3. The literature indicates that students evaluate instruction that they have not personally witnessed.
4. Limited research results indicate that some students will purposely falsify answers.

SURVEY

Of the four points listed above, the last has received the least empirical support. While Sproule (2000) and Stanfel (1995) gave actual cases of students falsifying responses, only one journal article was found that estimated the prevalence of giving purposeful misinformation among student populations (Clayson 2005b). Clayson (2005b), Sproule (2000), and Stanfel (1995) all found examples of students falsifying information on SETs, but objective measures of the extent of this practice are complicated by the manner in which the evaluations are utilized by both the students and the institutions. Much of the information requested on an SET is perceptual in nature, and students remain anonymous. Most evaluation systems place a high premium on confidentiality and nondisclosure. Consequently, most potential research of large samples would, of necessity, be based on student perception and self-disclosure rather than actual objective measures of misinformation.

A survey was conducted to answer the following questions:

1. What proportion of students will admit to giving purposeful misinformation on SETs?

2. What proportion of other students do they believe are also giving purposeful false information?

METHODOLOGY

Information was gathered from students enrolled in seven undergraduate marketing business core classes (principles of marketing and consumer behavior) at an AACSB (Association to Advance Collegiate Schools of Business)-accredited Midwestern university. These courses are required for a wide range of business majors as well as for public relations and some communication majors. In all, usable results were obtained from 219 students. In addition, as a type of validity check, a small sample of students ($n = 17$) was surveyed in the business college of an AACSB-accredited Southwestern university. Even though students at the Midwestern university put a higher value on SET usage than did the Southwestern students (99 percent of the Midwestern students thought that students should evaluate their teachers compared to 88 percent in the Southwest; $\chi^2 = 11.15$, degrees of freedom [df] = 1, $p = 0.001$), there were no statistically significant differences on any other of the research questions. Consequently, the two samples were combined for a total sample size of 236.

A copy of the questionnaire is in the Appendix. No student was identified by name or by student number. The researchers had no method of connecting any response with any student. The questionnaire asked the respondents what other students would do and what they had done. The questions were posed in both a positive and negative manner so that sequence bias would be minimized. With two exceptions, all the questions could be answered with a simple "yes" or "no."

Responses were elicited in the first ten minutes of class by the same researcher in all the classes. Responding students were not given a copy of the questionnaire. They had an answer sheet on which they could record their responses. The questionnaire was shown on an overhead projector one question at a time. The students could not read the questions in advance or reread questions already answered.

RESULTS

A summary of the results is in Table 1. The questions were rearranged by topic. Combining the responses, 56.2 percent (SE [sampling error] = 3.2 percent) of the students reported that they personally knew someone who gave a higher or lower evaluation than an instructor deserved. Almost one-third of the respondents reported giving an undeserved evaluation themselves (31.4 percent; SE = 3 per-

Table 1
Percentage Results

Question	Result
General Attitudes	
Do you think that students should evaluate their teachers?	Yes 98.3
How often do you think that teachers should be evaluated?	
a. Every class, every semester	67.4
b. In one class every semester	23.3
c. Once a year	7.2
d. Once every three years	1.3
e. Only if there is a problem	0.4
f. Never	0.4
Do you think that teachers who are evaluated become better teachers?	Yes 68.4
Do you think that administrators use evaluations when making teaching assignments?	Yes 44.4
Do you think that the comments written on the evaluations are read by the teacher?	Yes 74.2
Do you think that the comments written on the evaluations are read by administrators?	Yes 50.6
Do you think that the questions on the evaluation allow you to express what you really want to say on the evaluations?	Yes 55.1
If you had to give a score from 0 to 100 expressing how accurate these evaluations are in expressing how well an instructor teaches (for example, 20 would be 20% accurate, 90 would be 90% accurate, 100 would be 100% accurate), what score would you give?	Average = 62.9 (SE = 1.20)
Do you personally know of a teacher who was let go because of bad evaluations?	Yes 6.8
Misinformation	
Do you personally know a student who has given a teacher a higher than deserved evaluation for any reason?	Yes 35.3
Do you personally know a student who has given a teacher a lower than deserved evaluation for any reason?	Yes 48.7
Combined: higher or lower	Yes 56.2
Have you given a teacher a higher than deserved evaluation for any reason?	Yes 19.5
Have you given a teacher a lower than deserved evaluation for any reason?	Yes 19.9
Combined: higher or lower	Yes 31.4
Do you personally know of a student who wrote something untrue on the written comments because . . .	
a. they wanted to protect a teacher	Yes 5.4
b. they liked a teacher	Yes 17.8
Do you personally know of a student who wrote something untrue on the written comments because . . .	
a. they wanted to hurt a teacher	Yes 16.5
b. they disliked a teacher	Yes 36.5
Combined: positive or negative	Yes 41.7
Have you written something untrue on the written comments because . . .	
a. you wanted to protect a teacher	Yes 3.9
b. you liked a teacher	Yes 11.2
Have you written something untrue on the written comments because . . .	
a. you wanted to hurt a teacher	Yes 2.6
b. you disliked a teacher	Yes 12.9
Combined: positive or negative	Yes 19.4
Attitudes About Misinformation	
Do you think that giving an evaluation higher or lower than what a teacher deserves is a form of cheating?	Yes 43.2
Do you think that writing something untrue on an evaluation about a teacher is a form of cheating?	Yes 60.2
Estimate	
If you had to guess what percent of all evaluations submitted by students contained scores or information that the student knew was untrue, what percent of all evaluations do you believe that would be?	Average = 30.5 (SE = 1.60)

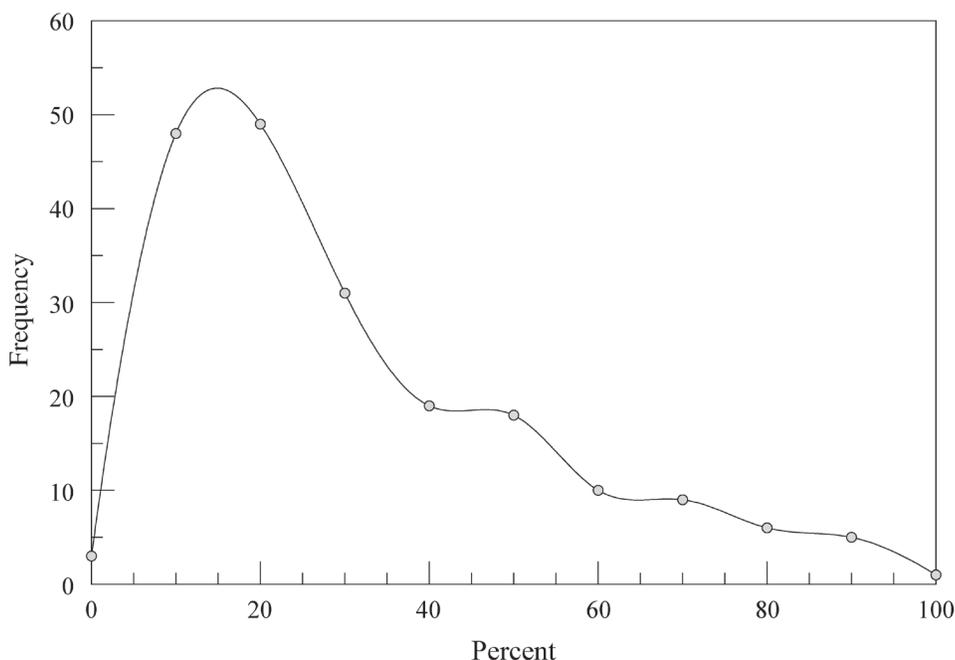
Note: SE = sampling error.

cent). On the SET instrument, students can voluntarily add written comments; 41.7 percent (SE = 3.2 percent) of the students reported that they personally knew someone who wrote something untrue on the evaluations, while

19.4 percent (SE = 2.6 percent) reported engaging in the same activity.

There were two specific questions asking whether misinformation on the evaluations was a form of cheating: "Do

Figure 1
Frequency of SET Perceived to Contain Purposeful Misinformation



you think that giving an evaluation higher or lower than what a teacher deserves is a form of cheating?" and "Do you think that writing something untrue on an evaluation about a teacher is a form of cheating?" The majority of the students (56.8 percent; SE = 3.2 percent) did not believe that giving an undeserved evaluation was a form of cheating. However, 60.2 percent (SE = 3.2 percent) did believe it was cheating to write something untrue on an evaluation. Forty-one percent of all the students thought both forms of misinformation were cheating, while 38 percent of the total sample did not think that either was a form of cheating.

The average estimate of the accuracy with which the evaluations expressed how well an instructor taught was 62.9 percent (SE = 1.20). Students estimated that 30.5 percent (SE = 1.60; median of all percent scores = 20) of all evaluations contained information that the responding student knew was not true (see Figure 1). There was a significant negative correlation between the two measures ($r = -0.321$, $p < 0.001$).

Female students were more likely to report misinformation as cheating and considered untruthful responses on both the structured ($\chi^2 = 3.88$, $df = 1$, $p = 0.049$) and written material ($\chi^2 = 16.22$, $df = 1$, $p < 0.001$) to be a form of cheating. Women believed that their comments were read more by instructors than did men ($\chi^2 = 9.18$, $df = 1$, $p = 0.002$), and they were less likely to believe that the evaluations allowed them to express what they wanted to say about the

class and instructor than did the men ($\chi^2 = 4.55$, $df = 1$, $p = 0.033$). There were no statistical differences in their estimates of misinformation.

DISCUSSION

The two survey questions can be addressed: (1) What proportion of students will admit to giving purposeful misinformation on SETs? Thirty-one percent of the respondents admitted to recording false information on the scale questions of the evaluations, while 19.4 percent admitted to adding untrue written comments on the evaluations. Combining both scales and written additions, in total, 37 percent of the students stated that they had submitted information in some form on the evaluations they knew were not deserved or were purposely false.

(2) What proportion of other students do they believe are also giving purposeful false information? The students estimated that about 56 percent of their fellow students have created evaluations they know are in error, and about 31 percent have written comments they know are untrue.

In total, the students estimated that about 30 percent of all evaluations contain material that the student who completed them knew was not true. Furthermore, the majority did not believe that this was a form of cheating as long as they were answering the structured questions. Just over 40 percent did not believe that adding untrue written

comments was a form of cheating. Two students added unsolicited comments to the questions about cheating. One wrote, "It is immoral, but not cheating"; another asked the question "How can it be 'cheating' because names aren't on them?"

Because the issue is student accuracy, it is also of interest to look at the errors the students made when completing the survey. Instructions were repeated three times, once before the first question was shown and twice as the questions were being answered by the students. Still, 11 percent of the students did not follow instructions on at least one aspect of the exercise. It is impossible to know for sure, but it is probable that the students took this assignment as seriously as the evaluations they routinely complete every semester.

LIMITATIONS

The majority of the students came from one university, and the proportions found may be larger or smaller at other institutions, but even though the sample size in the comparison university was small, the results from students in two universities in two different areas were found to be statistically similar. It is noteworthy that the Southwest sample did not support SET as much as those from the Midwest, but their responses about misinformation and dishonesty were statistically consistent. Furthermore, the results are compatible with the only other survey of student misinformation. Clayson (2005b) found in three different universities that 50 percent of the students admitted to giving higher or lower than average evaluations to an average teacher. In the current study, with much more detailed and explicit questions, the students reported that 56 percent of other students would engage in this behavior, but only 31 percent had done so themselves.

An interesting limitation is raised by the very nature of our research questions. If students give purposeful misinformation on SETs, would the current study also be untrustworthy? This creates a tautology that cannot easily be resolved because misinformation in this study would question the validity of the results but, at the same time, reinforce the contention that SET may be invalid for the same reason.

CONCLUSIONS

According to the literature review and the findings in this study, most students generally give honest responses and express heartfelt perceptions on the evaluations. This is reflected in how SET is utilized. In this lies a series of prob-

lems. The invalidity created by respondents giving false information is a pragmatic problem long recognized in the marketing literature (as an example, see "Interview with Susan Pogash" 1990). With SET, this creates three problems: first, using the evaluations as a guide for individual teaching improvement becomes compromised; second, making valid improvements in pedagogical issues and curriculum becomes more difficult. The third problem involves the dangers of making comparisons between instructors when lack of objectivity can potentially change the outcomes radically. Most SET procedures create highly skewed distributions that require an institution to use percentile rankings of instructors. Because of the nature of the data, small changes can create large effects. For example, in one of the current authors' college, 3 students in a class of 40 who change their evaluations from the extreme positive anchor to the extreme negative would reduce a ninetieth percentile instructor to the sixty-seventh percentile. The same students in a class of 20 would change the same ninetieth-percentile instructor to the forty-second percentile. If the students' estimate that about 31 percent of the evaluations contain misinformation is correct, then in a class of 30, 9 students changing their evaluation downward by only one unit on the SET scale would reduce an average instructor to a thirty-third-percentile teacher.

Even if a small percentage of students give inaccurate or dishonest evaluations, the resultant SET cannot be taken at face value. The evaluations must be interpreted in light of other evidence. An administrator utilizing the evaluations may learn something about students' opinions, perceptions, and attitudes toward instructors, but learn much less about the instructors' actual performance. This may be sufficient for some administrators, especially if they have adopted a customer model of students. Nevertheless, experience and the literature suggest that evaluations must be used with care and not mechanically applied as a measure of instructional merit. This is especially true when instructors are compared, and rewarded, using some sort of comparative scale created by SET outputs.

Proposed Solutions

The social science literature has been concerned about dishonest responses since the 1960s (Orne and Holland 1968). Researchers have recognized that there could be many reasons for dishonest responses, including a desire to compromise the results of whatever was being measured, a response that Maslow (1966) called the "screw you" effect." Popham (2001) suggests that dishonesty could be better controlled by making sure that responses were anonymous,

and by using class mean data rather than individual student data. Because most SET procedures already utilize both of these methods, it would seem that this is an inadequate solution.

Purposely falsifying data is considered a form of plagiarism (Park 2003). Yet the current study found that over 60 percent of students did not believe that giving purposeful misinformation on the structured questions of SET was cheating and 40 percent did not believe that adding written comments that were purposefully untrue was a form of cheating. Two solutions have been suggested to control cheating that may also have a positive effect on dishonest responses in SET. The institution, and professors individually, should indicate clearly that misreporting data on the evaluations is a form of cheating that entails penalties, and a request should be made that students not cheat (Nonis and Swift 1998). Business schools can also develop and publicly utilize an honor code (Eastman, Iyer, and Eastman 2006). This is particularly important because investigations of student dishonesty have found that cheating is a conscious decision made when the benefits outweigh the risks (Williams and Hosek 2003).

Future Research

Currently little is known about students furnishing incorrect and dishonest responses on SET. Before the problem identified in this study can be fully addressed, numerous questions need to be investigated:

1. The current study looked at willingness to produce purposeful misinformation, but what additional misinformation is simply due to error and misperception that may appear to be dishonest? This has been an issue that has long plagued pollsters (Lewis and Schneider 1982).
2. How are dishonest responses on SET related to other cheating? The percentage of students who cheat in class and on assignments has been estimated to be as low as 13 percent or as high as 95 percent, with rates between 60 to 75 percent considered common (Eastman, Iyer, and Eastman 2006). The percentage for marketing students may be even higher because some research has indicated that marketing students might cheat more than other business students (Chapman et al. 2004). Does dishonesty in SET follow the same pattern as class dishonesty, or does it follow a different paradigm?
3. Under what conditions would an instructor receive dishonest evaluations? The halo effect literature would suggest that if an instructor was liked or

disliked more than average, he or she could expect more dishonest responses. The grading literature would suggest that instructors who violated grading norms may find themselves more the target of dishonest evaluations.

4. Are there student characteristics that would predispose dishonest responses? In this survey, women were more likely to state that purposeful misinformation on SET was a form of cheating than did men. This is consistent with other research that indicates that female students are more ethically minded than male business students (Payan and Iyer 2006), but are attitudes consistent with behavior? There is some evidence that female students are more likely to cheat under low-risk conditions than males (Leming 1980). Currently, there is no risk in being dishonest on the evaluations.

Schools have made an extensive investment in the SET procedure in terms of both time and money. Surveys have indicated that almost 100 percent of business schools use some form of student evaluation of instruction. The authors' colleges are unusual only in the intensity of application. Both colleges have a policy of "Every faculty . . . evaluated by every class . . . every time." The process takes over 60 hours of direct class time in one college every semester, plus the time engaged by third parties (the faculty member cannot be present when the evaluations are given), plus the time to administer, handle, and evaluate the forms and store the results with required security. About every decade a committee meets for at least a year to evaluate the evaluations and propose changes. The adopted forms are printed, distributed, and computer scanned. The results are included in tenure, promotions, and merit pay considerations. Yet in this entire process, not one person, known to either author, has ever questioned the honesty of the responses or discussed what the consequences of dishonest responses would mean to the process.

Any issue that questions the usefulness and validity of SET needs to be carefully investigated. If not, then Stanley Fish (2005), who is a regular contributor to the *Chronicle of Higher Education*, must be taken seriously when he says of the SET process, "[It is] a whole lot of machinery with a very small and dubious yield."

REFERENCES

- Aiken, Lewis R. (1979), *Psychological Testing and Assessment*, Boston: Allyn and Bacon.
- Benz, Carolyn R., and Stephen J. Blatt (1996), "Meaning Underlying Student Ratings of Faculty," *Review of Higher Education*, 19 (4), 411-433.

- Carrell, Scott E., and James E. West (2010), "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors," *Journal of Political Economy*, 118 (3), 409-432.
- Cashin, William E. (1995), "Student Ratings of Teaching: The Research Revisited," Idea Paper no. 32, Center for Faculty Evaluation and Development, Division of Continuing Education, Kansas State University, Manhattan, KS.
- Chapman, Kenneth J., Richard Davis, Daniel Toy, and Lauren Wright (2004), "Academic Integrity in the Business School Environment: I'll Get By with a Little Help from My Friends," *Journal of Marketing Education*, 26 (3), 236-249.
- Clayson, Dennis E. (1989), "Halo Effects in Student Evaluations of Faculty: A Question of Validity," in *Marketing: Positioning for the 1990s*, Proceedings of the Southern Marketing Association, Robert L. King, ed., Atlanta: Southern Marketing Association, 271-275.
- (2005a), "Performance Overconfidence: Metacognitive Effects or Misplaced Student Experience," *Journal of Marketing Education*, 27 (2), 11-21.
- (2005b), "Within-Class Variability in Student-Teacher Evaluations: Example and Problems," *Decision Sciences Journal of Innovative Education*, 3 (1), 109-124.
- (2008), "A New Concept of Validity: Evaluation of Teaching, and the Production of Loincloths," in *Reaching New Heights in Marketing Education: Marketing Educators' Association 2008 Conference Proceedings*, Robert A. Lupton and Barbara L. Gross, eds., Madison, WI: Omnipress, 7-11.
- (2009), "Student Evaluations of Teaching: Are They Related to what Students Learn?" *Journal of Marketing Education*, 31 (1), 16-30.
- , and Debra A. Haley (2005), "Marketing Models in Education: Students as Customers, Products, or Partners," *Marketing Education Review*, 15 (1), 1-10.
- , and Mary Jane Sheffet (2006), "Personality and the Student Evaluation of Teaching," *Journal of Marketing Education*, 28 (2), 149-160.
- Comm, Clare L., and Dennis F.X. Manthaisel (1998), "Evaluating Teaching Effectiveness in America's Business Schools: Implications for Service Marketers," *Journal of Professional Service Marketing*, 16 (2), 163-170.
- Cruse, Daniel B. (1987), "Student Evaluations of the University Professor: Caveat Professor," *Higher Education*, 16 (6), 723-737.
- Dershowitz, Alan M. (1994), *Contrary to Popular Opinion*, New York: Berkley Books.
- Dunegan, Kenneth J., and Mary W. Hrivnak (2003), "Characteristics of Mindless Teaching Evaluations and the Moderating Effects of Image Compatibility," *Journal of Management Education*, 27 (3), 280-302.
- Eastman, Jacqueline K., Rajesh Iyer, and Kevin L. Eastman (2006), "Addressing Academic Dishonesty: The Implications for Business Schools, Professors, and Students," *Journal for Advancement of Marketing Education*, 9 (Winter), 1-8.
- Engelland, Brian T. (2004), "Making Effective Use of Student Evaluations to Improve Teaching Performance," *Journal of Advancement of Marketing Education*, 5 (Winter), 38-46.
- Felton, James, Peter T. Koper, John Mitchell, and Michael Stinson (2004), "Web-Based Student Evaluations of Professors: The Relations Between Perceived Quality, Easiness and Sexiness," *Assessment & Evaluation in Higher Education*, 29 (1), 91-108.
- Fish, Stanley (2005), "Who's In Charge Here?" *Chronicle of Higher Education*, February 4 (available at <http://chronicle.com/article/Whos-In-Charge-Here-/45097/>).
- Follman, John (1984), "Pedagogue . . . Paragon and Pariah . . . 20% of the Time: Implications for Teacher Merit Pay," *American Psychologist*, 39 (9), 1069-1070.
- Gray, Mary, and Barbara R. Bermann (2003), "Student Teaching Evaluations: Inaccurate, Demeaning, Misused," *Academe*, 89 (5), 44-46.
- Gremler, Dwayne D., and Michael A. McCollough (2002), "Student Satisfaction Guarantees: An Empirical Examination of Attitudes, Antecedents, and Consequences," *Journal of Marketing Education*, 24 (2), 150-160.
- Grimes, Paul W. (2002), "The Overconfident Principles of Economics Students: An Examination of Metacognitive Skill," *Journal of Economic Education*, 33 (1), 15-30.
- , Meghan J. Millea, and Thomas W. Woodruff (2004), "Grades—Who's to Blame? Student Evaluation of Teaching and Locus of Control," *Journal of Economic Education*, 35 (2), 129-147.
- Haskell, Robert E. (1997), "Academic Freedom, Tenure, and Student Evaluation of Faculty: Galloping Polls in the 21st Century," *Education Policy Analysis Archives*, 5 (6) (available at <http://epaa.asu.edu/epaa/v5n6.html>).
- "Interview with Susan Pogash: On the Problem of Professional Respondents" (1990), *Applied Marketing Research*, 30 (3) (available at www.consumerfocus.com/interview.html).
- Johnson, Valen E. (2003), *Grade Inflation: A Crisis in College Education*, New York: Springer.
- Kelly-Woessner, April, and Matthew C. Woessner (2006), "My Professor Is a Partisan Hack: How Perceptions of a Professor's Political Views Affect Student Course Evaluations," *PS: Political Science & Politics*, 39 (3), 495-501.
- Kennedy, Ellen J., Leigh Lawton, and E. Leroy Plumlee (2002), "Blissful Ignorance: The Problem of Unrecognized Incompetence and Academic Performance," *Journal of Marketing Education*, 24 (3), 243-252.
- Leming, James (1980), "Cheating Behavior, Subject Variables, and Components of the Internal-External Scale Under the Low Risk Conditions," *Journal of Educational Research*, 74 (2), 83-87.
- Lewis, I.A., and William Schneider (1982), "Is the Public Lying to the Pollster? A Question of Truth," *Public Opinion*, 5 (2), 42-47.
- Machina, Kenton (1987), "Evaluating Student Evaluations," *Academe*, 73 (May-June), 19-22.
- Marks, Ronald B. (2000), "Determinants of Student Evaluations of Global Measures of Instructor and Course Value," *Journal of Marketing Education*, 22 (2), 108-119.
- Marsh, Herbert W., and Lawrence A. Roche (2000), "Effects of Grading Leniency and Low Workload on Students' Evaluations of Teaching: Popular Myth, Bias, Validity, or Innocent Bystanders?" *Journal of Educational Psychology*, 92 (1), 202-228.
- Maslow, Abraham R. (1966), *Psychological Testing and Assessment*, Boston: Allyn and Bacon.
- McCroskey, James C., and Thomas J. Young (1981), "Ethos and Credibility. The Construct and Its Measurement After Three Decades," *Central States Speech Journal*, 32 (1), 24-34.
- Moreland, Richard, John Miller, and Frances Laucka (1981), "Academic Achievement and Self-Evaluations of Academic

- Performances," *Journal of Educational Psychology*, 73 (3), 335-344.
- Nonis, Sarah A., and Cathy O. Swift (1998), "Deterring Cheating Behavior in the Marketing Classroom: An Analysis of the Effects of Demographics, Attitudes, and In-Class Deterrent Strategies," *Journal of Marketing Education*, 20 (3), 188-199.
- Orne, Martin T., and Charles H. Holland (1968), "On the Ecological Validity of Laboratory Deceptions," *International Journal of Psychiatry*, 6, 282-293.
- Orsini, Joseph L. (1988), "Halo Effects in Student Evaluations of Faculty: A Case Application," *Journal of Marketing Education*, 10 (2), 38-45.
- Park, Chris (2003), "In Other (People's) Words: Plagiarism by University Students: Literature and Lessons," *Assessment & Evaluation in Higher Education*, 28 (5), 471-488.
- Payan, Janice M., and Vish Iyer (2006), "Are Female Marketing Students Always More Ethical Than Male Marketing Students?" *Journal for Advancement of Marketing Education*, 9 (Winter), 26-30.
- Popham, W. James (2001), *The Truth About Testing: An Educator's Call to Action*, Alexandria, VA: Association for Supervision and Curriculum Development.
- Remmers, Herman H., and G.C. Brandenburg (1927), "Experimental Data on the Purdue Rating Scale for Instruction," *Educational Administration and Supervision*, 13, 519-527.
- Reynolds, David V. (1977), "Students Who Haven't Seen a Film on Sexuality and Communication Prefer It to a Lecture on the History of Psychology They Haven't Heard: Some Implications for the University," *Teaching of Psychology*, 4 (2), 82-83.
- Seldin, Peter (1993), "The Use and Abuse of Student Ratings of Professors," *Chronicle of Higher Education*, 21 (July), A40.
- (1999), *Changing Practices in Evaluating Teaching: A Practical Guide to Improving Faculty Performance and Promotion/Tenure Decisions*, Bolton, MA: Anker.
- Sproule, Robert (2000), "Student Evaluation of Teaching: A Methodological Critique of Conventional Practices," *Educational Policy Analysis Archives*, 8 (50) (available at <http://epaa.asu.edu/ojs/article/view/441/564/>).
- Stanfel, Larry E. (1995), "Measuring the Accuracy of Student Evaluations of Teaching," *Journal of Instructional Psychology*, 22 (2), 117-125.
- Tang, Thomas L., and Theresa L. Tang (1987), "A Correlation Study of Students' Evaluations of Faculty Performance and Their Self-Ratings in an Instructional Setting," *College Student Journal*, 21 (Spring), 90-97.
- Weinberg, Bruce A., Belton M. Fleisher, and Masanori Hashimoto (2007), "Evaluating Methods for Evaluating Instruction: The Case of Higher Education," Working Paper no. 12844, National Bureau of Economic Research, Cambridge, MA, January (available at www.nber.org/papers/w12844/).
- Williams, Melanie S., and William R. Hosek (2003), "Strategies for Reducing Academic Dishonesty," *Journal of Legal Studies Education*, 21 (1), 87-107.
- Wilson, Robin (1998), "New Research Casts Doubt on Value of Student Evaluations of Professors," *Chronicle of Higher Education*, 44 (19), A12-A14.

APPENDIX

Questionnaire

Following are some questions about student teacher evaluations. Please be very honest. The information will be used.

1. Do you think that students should evaluate their teachers?
2. How often do you think that teachers should be evaluated?
 - a. Every class, every semester
 - b. In one class every semester
 - c. Once a year
 - d. Once every three years
 - e. Only if there is a problem
 - f. Never
3. Do you think that teachers who are evaluated become better teachers?
4. Do you think that administrators use evaluations when making teaching assignments?
5. Do you personally know of a teacher that was let go because of bad evaluations?
6. Do you personally know a student who has given a teacher a higher than deserved evaluation for any reason?
7. Do you personally know a student who has given a teacher a lower than deserved evaluation for any reason?
8. Have you given a teacher a higher than deserved evaluation for any reason?
9. Have you given a teacher a lower than deserved evaluation for any reason?
10. Do you think that the comments written on the evaluations are read by the teacher?
11. Do you think that the comments written on the evaluations are read by administrators?
12. Do you personally know of a student who wrote something untrue on the written comments because:
 - a. they wanted to protect a teacher
 - b. they liked a teacher
13. Do you personally know of a student who wrote something untrue on the written comments because:
 - a. they wanted to hurt a teacher
 - b. they disliked a teacher
14. Have you written something untrue on the written comments because:
 - a. you wanted to protect a teacher
 - b. you liked a teacher
15. Have you written something untrue on the written comments because:
 - a. you wanted to hurt a teacher
 - b. you disliked a teacher
16. Do you think that the questions on the evaluation allow you to express what you really want to say on the evaluations?
17. Do you think that giving an evaluation higher or lower than what a teacher deserves is a form of cheating?
18. Do you think that writing something untrue on an evaluation about a teacher is a form of cheating?
19. If you had to give a score from 0 to 100 expressing how accurate these evaluations are in expressing how well an instructor teaches (for example, 20 would be 20% accurate, 90 would be 90% accurate, a 100 would be 100% accurate), what score would you give?
20. If you had to guess what percent of all evaluations submitted by students contained scores or information that the student knew was untrue, what percent of all evaluations do you believe that would be?
21. I am: M F

Fold the paper and turn it in.