Personality and the Student Evaluation of Teaching

Dennis E. Clayson and Mary Jane Sheffet

Students' perception of the instructor's personality and the evaluation of instruction were found to be strongly related. Students appear to be using a contaminated measure to establish personality and its relationship to the evaluations. The findings do not support the contention that the association reflects a valid measure of instruction. Implications for changing evaluations and the use of the instruments are discussed.

Keywords: SET; student evaluation of teaching; personality; evaluation of personality; marketing education

Because almost all learning institutions use some sort of evaluation of instruction, and because these have the potential of establishing reputations, merit pay, and promotion, it would make sense that a vigorous debate would arise over the validity of the process. Yet after 40 years of investigation and thousands of published articles, instead of a converging consensus, researchers have aligned themselves into divergent camps. Researchers and writers from the colleges of education generally have defended student evaluations of teaching (SETs), while those from other disciplines, especially business, have questioned their validity. Marsh and Roche (2000, p. 202), who champion the validity of SET, stated that their studies "debunk popular myths" and prefaced their introduction with a running head that read, "Popular Myths: An Anecdotal Approach to Bias." On the other hand, Johnson (2003), a biostatistician, maintains that SETs "are not reliable indicators of teaching effectiveness" (p. 237) and strongly suggests that the SET supporters actually may be defending ideology and vested instruments more than scientific rigor.

Much of the debate has centered on the grade-and-evaluation relationship (see Clayson 2004; Clayson, Frost, and Sheffet 2006; Johnson 2003; Marsh and Roche 1999, 2000; and Stumpf and Freedman 1979 for extensive reviews). At the same time, another area of disagreement has generated a surprisingly low level of published research, given its centrality to the validity issue. Put simply, does the students' perception of the instructor's personality influence their evaluation of instruction? If so, troubling questions would be raised about the evaluation process. For example, can crowd-pleasing professors get good evaluations irrespective of student learning and achievement? On a more subtle note, because personality generally is considered to be an intrinsic, personal, and long-lasting set of characteristics, is it possible to make long-term changes in an individual instructor's evaluations? Can traits that would raise evaluations be taught? If not, should teaching be thought of as a vocation, and if so, what is the purpose of a college of education? In graduate training, should precious time be spent in preparing future PhDs to teach?

Purpose of Research

This study adds to the debate by looking at how the students' perception of instructors' personality measures are related to the evaluation of instruction. Students were asked to evaluate their instructor's personality and their perception of the instructor and class (SET) at four different times during the course of a term. The study compares these personality measures, taken through the entire duration of the term, with the end-of-term evaluations. To control for environmental, student, faculty, and interactive influences, the change in the perception of personality during the last six weeks of the term were compared to the change in evaluations made during the same time period. We know of no study that has used these controls.

Literature Review

Defenders of the SET process generally deny that personality has any influence on the evaluations that could invalidate the process. Cashin (1995, p. 4) states that personality is one of many variables "not related to student ratings." Felder (1995) refers to the idea that SETs are popularity contests as a myth. Boice (1992) is unequivocal

Dennis E. Clayson is a professor of marketing in the College of Business Administration, University of Northern Iowa. Mary Jane Sheffet is currently professor and head of the Department of Marketing, College of Business Administration, University of Northern Iowa.

in stating, "In fact, the gist of research is that measures of personality and popularity correlate at low, usually insignificant levels with SETs" (p. 2). Theall and Franklin (2001), who are careful to write of personality only in terms of popularity, state, "There is no basis for this argument and no research to substantiate it" (p. 49). In general, researchers publishing from within educational disciplines report finding few personality traits that correlate with student ratings (Braskamp and Ory 1994; Centra 1993).

These reports, however, generally are derived from summaries that may have skimmed over a more detailed look at the data. For example, Cashin (1995) cites Feldman (1986) in stating that there is no relationship between personality traits and SET. This finding is reported under a section looking at instructor variables. No mention of personality is made in reviews of student variables that may be related to SET. It is true that Feldman's (1986) careful review of the personality-evaluation literature did find that the instructors' own view of their personality showed little correlation with SET, but he also found that the students' perception of the instructors' personality traits was strongly related to SET. Feldman (1986) states, "Not only are a wide variety of perceived traits [personality traits] associated with rated effectiveness of teachers, but the relationships tend to be moderate to strong" (p. 157).

Feldman's contention is consistent with findings from a wide variety of sources. Students appear to form their opinions of a class and an instructor very early in a course, and some evidence indicates that subsequent class and learning experiences may do little to change that opinion (Hewett, Chastain, and Thurber 1988; Ortinau and Bush 1987; Sauber and Ludlow 1988). Further, studies that experimentally have manipulated classroom conditions have found interesting effects of the instructors' perceived personalities on the evaluations (Naftulin, Ware, and Donnelly 1973). As an example, Widmeyer and Loy (1988) conducted an experiment in which all students were exposed to the same guest instructor, but half received prior descriptions of the instructor indicating that he was warm and the other half that he was cold. Not only did the students in the warm group rate the instructor higher on positive aspects of personality, they also rated the warm instructor as having better teaching ability. Harvard psychologists (Ambady and Rosenthal 1993) investigated students' reactions to randomly selected 30-second clips of soundless videotapes of actual classroom instruction and found the clips highly correlated with end-of-course evaluations. Evaluations based on 30-second exposures were no more significant than judgments based on 6-second clips. Personality traits identified by independent raters were highly correlated with the evaluations. These traits, in order, were *optimistic, confident, dominant, active, enthusiastic,* and *likeable.*

Several studies have found large associations between personality variables and the evaluation outcomes. Murray (1975) reported that colleague ratings of instructor personality accounted for 67% of the variance of between-instructor student ratings. Sherman and Blackburn (1975) found that 77% of the variance of the evaluations could be explained by the personal characteristics of the instructors. The relationship was so high that they concluded, "A professor wishing to improve his perceived effectiveness may best begin on personal attributes rather than focus his energy on course functions and activities which, on the surface, seem more readily open to alteration" (p. 130). Other researchers in a variety of settings, including business students, have found similar large effects (Erdle, Murray, and Rushton 1985; Marks 2000; Murray, Rushton, and Paunonen 1990).

One structural modeling study of marketing students found that the total effect of personality on the student evaluation of faculty was very high, with each standard-deviation change in personality resulting in a 0.83 standard-deviation change in the evaluations. Personality was found to be significantly related to every other factor in the study, including the students' perception of the instructor's knowledge and fairness. It was negatively related to rigor and positively related to the students' perception of how much they had learned (Clayson and Haley 1990). The researchers concluded that the evaluations essentially constitute a likeability scale. A more recent study found the same effects (Marks 2000).

Consistent with a personality interpretation, the evaluations have been found to be remarkably consistent for instructors, even during periods as long as 13 years. Previous teaching experience was not related to this consistency (Marsh and Hocevar 1991). Because most professionals improve their performance with constant practice, what could the evaluations be measuring that would not change? Business students were asked to evaluate instructors on a number of characteristics and were asked how these terms would change over time. Attributes such as knowledge, fairness, and organization were perceived as improving through time with experience. The students' perceptions of instructor characteristics described as *responsive, interesting, cares, stimulating,* and *open* remained constant. The student response indicated that the evaluations are heavily biased toward personality variables and are less influenced by the instructor's perceived knowledge, fairness, or even the perception of students' own learning (Clayson 1999).

Some writers in marketing education simply assume, a priori, a relationship between effective teaching and personality characteristics. Lantos (1997) encourages instructors to use humor, fun and games, learning students' names, and being genuine as methods of motivating students. After reviewing the literature and conducting their own study, Foote, Harmon, and Mayo (2003) concluded, "those [instructors] who score highly on evaluations may do so not because they teach well, but simply because they get along well with students" (p. 17).

The importance of a possible association between personality and evaluation goes beyond potential problems with popularity and crowd pleasing. Kulik (2001) is typical of those who defend SET when he makes three claims that he

says demonstrate the validity of the instruments: (1) student ratings agree with student comments in interviews, (2) student ratings agree with observer ratings, and (3) student ratings agree with alumni ratings. If it is shown that personality constitutes a large proportion of SET, one would expect to find these associations irrespective of what else SET happens to be measuring. These associations' being present (or absent) would be moot in any discussion of validity.

Research Questions

We must first establish that an association between personality and evaluation exists within our data. Specifically, (1) does a relationship exist between personality characteristics and the evaluations in marketing and business core classes?

If the relationship is found to exist, then we would wish to determine when it is established. Specifically, (2) how early in the term does it develop?

If the students' perception of personality changes during the term and there is a corresponding change in the evaluations, then the students' perception of personality cannot be associated with any long-term classroom effects or relatively permanent teaching competencies. Consequently, (3) will changes in personality after a class is well established be related to changes in the evaluations given?

If question three is answered in the affirmative, another issue would be raised. There are many definitions of personality, but since the time of Allport (1955), there has been a general consensus stretching across the spectrum from biophysical to personal-essence definitions that the traits or behaviors that define personality are relatively enduring or consistent. Consequently, if personality is perceived as changing in the last weeks of a 16-week term, with what is that change associated? The answer to this question will help define how students are using personality in the evaluation process. Feldman (1986), for example, suggests that students may be comingling several other factors with personality so that the effects are contaminated. Clayson and Haley (1990) proposed that SET essentially creates a likeability scale. Orsini (1988) found a strong halo effect. Consequently, another question can be investigated: (4) how much of the change in personality, resulting in a change in the evaluations, is related to other factors, such as general likeability and halo effects?

Personality Measures

It became obvious to psychological researchers in the last century that personality was multidimensional. The number of dimensions, or traits, however, was debated vigorously. Raymond Cattell suggested 16 factors of personality, while Osgood suggested only three (Osgood, Suci, and Tannenbaum 1957). There has been a general consensus reached in the last 20 years that an "adequate taxonomy for personality attributes" could be created by five factors (Digman 1990, p. 418). These have been referred to as the Big Five, or as the Five-Factor Model of personality. The five types are dimensions of personality, not types of people. The factors have been found to be very stable over long periods of a person's life (Soldz and Vaillant 1999) and are largely heritable, that is, genetic (Jang et al. 1998). They seem to be unrelated to culture in that they have been found in societies as diverse as those in Germany and China (McCrae and Costa 1997).

The five factors are the following: (1) Agreeableness: People with a high score tend to be friendly, trusting, and cooperative. Persons with a low score are often more aggressive and less cooperative; (2) Conscientiousness: This trait indicates how organized and persistent a person is in pursuing goals. Persons with a high score tend to be methodical, well organized, and respectful of their duties. Those with a low score are less careful and focused and more likely to be distracted; (3) Stability: Persons with a low score on this trait tend to be prone to insecurity and emotional distress. Those with a high score are more relaxed, less emotional, and less prone to distress; (4) Extroversion: An extroverted person will seek out the company of others and be energized by such interactions. Persons with a low score tend to be more quiet and reserved; and (5) Creativity (Openness): Persons with a high score on this trait tend to be open minded, creative, and interested in culture. Those with a low score are more down to earth and more practical in nature.

METHOD

Procedure

Data for this study were obtained from a database created at our university during the spring semester of 2003. Only the variables pertinent to this study are outlined below. Nine instructors from 14 sections of introductory, undergraduate business courses (six sections of Organizational Management and eight sections of Principles of Marketing) gave permission for the study to be conducted in their classes. These courses are required for all business students.

Week Zero

On the first meeting of the class, the instructors introduced themselves, turned the class over to a researcher, and left the room. At this point, no syllabus had been distributed and each class had had fewer than 5 minutes of exposure to the instructor. All students were given a consent form stating that if they agreed to be a subject, one of the researchers would access their cumulative grade point average (GPA) and their final grade in the current class at the end of the semester. All data would be collected by student identification only. The instructor never would see any individual's data or any information that would allow any student to be identified. Students could withdraw from the study at any time. The researchers taught none of the sections studied.

Students who signed the consent forms then were asked to complete a questionnaire that contained all the variables outlined below plus a set of demographic questions.

Subsequent Sessions

The class sections were evaluated again after one week, then at Week 10, and again at Week 16 of a 16-week term. At Week 10, the students had had exposure to the instructor and her or his grading standards for 10 weeks, and all students had access to their midterm grades. The questionnaires at each subsequent session were identical to the one given before the class began, except that no demographic data were gathered.

Participants

In total, the database consisted of 727 students. Some students dropped out of class or out of the study, not all students attended every class, and not all questions on every questionnaire were answered by each student. Consequently, the sample size for any given variable could vary from the total. Data used by this study came from 498 students who completed all the pertinent information for both Week 10 and Week 16. Seventeen of these students were enrolled in two of the classes used. Their data were removed from the larger of the two classes, leaving a final sample size of 481.

Of these students, 50% were female, 85% were juniors or seniors, and 15% were sophomores. The average age was 20.9 years $(SD = 1.87)$, and the average cumulative GPA was 3.06 $(SD = 0.43)$ at the beginning of the study.

Variables

Evaluation (SET)

SET was measured by using the five statements from the respondents' university evaluation (*the instructor created an atmosphere conducive of learning, the instructor explains material appropriately, the instructor shows interest in student learning, the instructor sets high but reasonable standards*, and *rate your satisfaction with your learning in this class*). These were summed and averaged. Each statement could be answered with a letter grade (A through F). Cronbach's alpha was .913 for Week 16 and 0.895 for Week 10. A second unambiguous SET measure, "What grade would you give your instructor?" from each testing period also was identified. The measures were similar. The correlations between the two were $r = .886$ $(r^2 = .78)$ at Week 16 and $r = .885$ $(r^2 = .78)$ at Week 10. Consequently, the two measures of evaluation were summed to create a total evaluation measure (*Eval*) used as the dependent variable in the study (Cronbach's alpha was .929 for Week 16 and .911 for Week 10). This measure is now similar to the dependent variables used in most of the studies reviewed by Feldman (1986).

The second dependent variable was the summed evaluations at Week 16 minus the same measure at Week 10.

Evaluation Difference = Evaluation (Week 16)
— Evaluation (Week 10).

Personality Measures

Since Big Five personality evaluations can be long and complicated, the five factors of the personality model were measured by using a simple semantic scaling device. The question read, "From what you know now, rate this instructor on the following dimensions":

Disagreeable: 1 2 3 4 5 6 7 : *Agreeable*
Not conscientious: 1 2 3 4 5 6 7 : *Conscientious*
Emotionally unstable: 1 2 3 4 5 6 7 : *Emotionally stable*
Introverted: 1 2 3 4 5 6 7 : *Extroverted*
Unimaginative–
uncreative: 1 2 3 4 5 6 7 : *Imaginative–creative*

As a validity check of the procedure, 72 students who did not participate in the initial study but who were in similar classes were asked to complete the current survey instrument along with an established Big Five model inventory already in use. The complete personality inventory contains 41 items and has known reliability and validity measures (see Buchanan 2001 for a detailed review of validity issues and for references). The two instruments were presented in a counterbalanced fashion. The current instrument was found to have both concurrent and predictive validity. The short form correlated $r = .814$ on a global personality measure with the larger inventory, and a LISREL confirmatory factor analysis showed a standardized associational coefficient between the short form and the inventory of .97. Furthermore, there were no significant differences between associational measures on the study form and the personality inventory with any of the major variables of the study.

Although the five factors represent separate personality characteristics, a high response to four of the measures can be seen as positive in the evaluation of instruction, and previous research has shown that extroversion in the introverted-extroverted scale is seen as a positive teaching attribute (Erdle, Murray, and Rushton 1985). Consequently, the five factors can be summed and averaged to produce a compensatory, global measure of the overall negative-positive perception of personality. This variable simply was called global personality positive, or GPP. Global personality positive is not personality in that the construct usually is defined as a cluster of independent traits or characteristics. Nevertheless, a student could, for example, believe an instructor was positive on one or several factors but not on all and still perceive the instructor as having a good or a bad personality globally and independently of the perception of any specific factor. Further, many previous studies did not use a personality inventory when measuring personality but instead relied on some global measure. Consequently, when studying SET, we believe that the addition of the GPP measure was an important distinction that needed to be included within any comprehensive statistical look at a possible relationship between personality and evaluation.

At each time period, Cronbach's alpha for GPP was the following: Week 0: GPP0, alpha = .909; Week 1: GPP1, alpha = .814; Week 10, GPP10, alpha = .785; Week 16, GPP16, alpha = .836. To create the difference scores, each measure at Week 10 was subtracted from the corresponding score at Week 16.

Finally, another scale was added to investigate Clayson and Haley's (1990) claim that the evaluations were essentially a likeability scale.

Unlikable: 1 2 3 4 5 6 7 : *Likable*

It is important to note that although likeability and personality are related, they are not the same construct. Although one may color the other, it is still true that a person's personality traits can be determined independently of any observer's global perception of general likeability. Extroversion, for example, can be recognized in others irrespective of whether the observer likes or dislikes people who are extroverted. At the same time, a student observer may be cognizant that an instructor has a compensatory positive personality and still not like the instructor.

Other Measures

Because expected grades have been found to be associated with the evaluations, at every testing period the students were asked to estimate what grade they would receive for the class. The difference score was simply

Expected Grade Difference = Expected Grade (Week 16)
– Expected Grade (Week 10).

Student characteristics were measured with four variables. *GPA* is the cumulative student GPA at the beginning of the course. *Age* is the age of the student at the beginning of the term, and *Sex* is a dummy variable with 0 = male and 1 = female. *Class performance* was measured by the actual grade received for the course. Student expectations of the class were measured at Week 0 by asking the students what grade they expected to receive in the class at the end of the term. Also at Week 0, each student could indicate whether he or she had previous information about how rigorous the instructor was as a grader; this dummy variable was simply labeled as *heard* (0 = had not heard something about the instructor and 1 = had heard about the instructor). Sophomores were allowed to take these courses only if they had previously met specified conditions. This was the first business class within the primary core that these students had taken. A preliminary look at the data set had shown that they rated the instructors differently than did the more experienced juniors and seniors. Consequently, this variable was added as a dummy (0 = juniors and seniors and 1 = sophomores). At each testing period, the students were asked the following questions: "Compared to other classes you have had, would you consider this class to be . . . ?" and "Compared to other instructors you have had, would you consider this professor to be . . . ?" Both questions could be answered by *easy*, *average*, or *hard*. A preliminary investigation of the data set had shown that average conditions received the highest SET. Consequently, these variables were added as dummies (0 = not average and 1 = average).

Both Orsini (1988) and Clayson (1989) found that a halo effect was related to the evaluations by marketing students. In both cases, the halo was measured by the lack of variation in the measures. Students who rated an instructor highly had a tendency to rate all questions or statements in the evaluation instrument highly and with little variation. When students wished to give a low evaluation, they seemed to be searching through the questions or statements more carefully, looking for a way to express why they were dissatisfied, which results in larger variation in the evaluation instruments. Since this questionnaire had five scales for personality and five scales with the college's evaluation instrument, *halo* was operationally defined as the average difference in these measures. Specifically, halo was measured as

$$-\text{Halo} = \left(\sum_{i=1}^{5} \sum_{j=i+1}^{5} (P_i - P_j) + \sum_{i=1}^{5} \sum_{j=i+1}^{5} (Ev_i - Ev_j) \right) / 2. \quad (1)$$

RESULTS

Pretest of Variables

It is possible that the sample used in this study was biased through repeated measures or by systematic factors created by some uniqueness of those who dropped out of the study. Consequently, all the variables used in the database were compared between three groups: all measures of the variables present in four administrations, three administrations, and two administrations (the last measure had to be present). The ANOVA revealed no significant differences based on how often the students responded to the same instructor. Because the sample used consists only of those students who completed the needed information from both Week 10 and Week 16, there may be systematic differences between the sample used ($n = 481$) and the total sample base ($n = 727$). The variables used in the study were compared with the same variables in the total sample. None of the differences were significant except for *Age*. The students used in the present study were slightly younger than the total sample by slightly more than 2 months. As will be shown below, age had no significant association with the dependent variables of the study. It appeared as if pretesting effects, missing data, and/or students' dropping out of the study did not influence or compromise the findings.

Tests of Research Questions

1) Does a relationship exist between personality characteristics and the evaluations in marketing and business core classes? The Pearson's correlation between variables is given

TABLE 1
CORRELATION MATRIX OF ALL STUDY VARIABLES AT END OF TERM

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Evaluation	—	.803*	.315*	.381*	.060	.056	.031	−.109	.073	.292*	.220*	.025	.150*
2. GPP		—	.250*	.357*	.001	.050	.060	−.112	.076	.178*	.133*	.014	.108*
Conscientiousness	.739*	.843*	.201*	.248*	.028	.034	.047	−.144*	.038	.226*	.127*	−.037	.090
Creativity	689*	.803*	.197*	.355*	.006	.050	.040	−.036	.141*	.125*	.105	.004	.073
Agreeableness	.717*	.829*	.246*	.283*	.014	.012	.026	−.157*	.052	.215*	.149*	.016	.106
Stability	.635*	.802*	.168*	.240*	.016	.018	.036	−.078	.090	.137*	.119*	−.028	.042
Extroversion	.347*	.613*	.179*	.259*	.006	.114*	.059	−.054	−.009	.030	.018	.123	.143*
3. Expected grade			—	.073	.006	.393*	.047	.056	.062	.182*	.176*	.300*	.648*
4. Halo effect				—	.095	−.025	.057	−.042	.022	.081	.120*	−.055	.000
5. Age					—	−.144*	−.118*	−.241*	.070	.007	.079	−.086	−.061
6. GPA						—	.196*	.160*	−.026	−.082	−.051	.423*	.568*
7. Sex[a]							—	.062	.081	−.041	.066	−.031	.094
8. Sophomore[a]								—	.122	−.164*	−.058	.021	.088
9. Heard[a]									—	.031	.001	.012	044
10. Hard class[a]										—	.320*	−.045	.081
11. Hard grader[a]											—	.019	.058
12. Initial expectations												—	.289*
13. Final grade													—

NOTE: GPP = global personality positive; GPA = grade-point average.
a. Point biserial correlations: sex (0 = male, 1 = female), sophomore (0 = junior or senior, 1 = sophomore), heard (0 = had not heard, 1 = had heard), hard class (0 = not average, 1 = average), hard grader (0 = not average, 1 = average).
* Significant at $p < .01$.

in Table 1. By the last week of the term, GPP accounted for over 64% of the variance ($r = .803$) in the evaluation measure. The five personality factors independently account for between 12% and 55% of the variance in the evaluations ($r = .347$ for extroversion to $r = .739$ for conscientiousness). A regression of the five personality dimensions shows that each was a significant predictor of the final evaluation except for extroversion (see Table 2). The five factors accounted for almost 70% of the variance in the evaluation measure.

As shown in Table 3, controlling for expected grades, halo effects, student characteristics such as gender, age, GPA, and year in school, the perceptions of the difficulty of the course and instructor, the initial grade expectations, and the class performance as measured by the grade given at the end of the term did not reduce GPP and likeability to insignificance.

2) If an association does exist between personality and evaluation, how early in the term does it develop? After fewer than 5 minutes of exposure to the instructor, the initial ratings of GPP are significantly correlated with the final evaluation given 16 weeks later (GPP, $r = .148$, $t (406) = 3.07$, $p = .002$). As shown in Table 4 and Figure 1, the association between personality measures and the final evaluation grows stronger at every subsequent testing period. Each of the five factors of personality (except extroversion) at Week 0 also were significantly related to the final evaluation at Week 16 (conscientiousness, $r = .128$, $p < .01$; creativity, $r = .162$, $p < .01$; agreeableness, $r = .142$, $p < .01$; stability, $r = .142$, $p < .01$; and extroversion, $r = .076$, $p = .123$).

TABLE 2
LINEAR REGRESSIONS: EVALUATION (SET) BY PERSONALITY VARIABLES

	B	Beta	t	Sig	Tolerance
Conscientiousness	0.41	.330	8.16	.000	0.41
Creativity	0.34	.296	8.53	.000	0.56
Agreeableness	0.31	.237	6.00	.000	0.43
Stability	0.15	.118	3.24	.001	0.51
Extroversion	−0.01	−.005	−0.17	.866	0.81
Constant	−0.77		−3.26	.001	

NOTE: Variables arranged by the magnitude of *beta*. SET = student evaluation of teaching.
$r^2 = .687$, $F (5,467) = 207.70$, $p < .0001$

3) Will changes in personality after a class is well established be related to changes in the evaluations given? Changes in the GPP between the 10th and 16th weeks of the term were related significantly to changes in the evaluation measured during the same time period ($r = .542$, $r^2 = .294$). Changes in each of the five factors individually were correlated significantly with the change in the evaluations (conscientiousness, $r = .362$; creativity, $r = .381$; agreeableness, $r = .458$; stability, $r = .319$; and extroversion, $r = .179$, with $p < .01$). Controlling for the change in halo effect, student characteristics, the students' perception of the difficulty of the class, and perceived and actual grades did not reduce the relationship between GPP change and evaluation to insignificance (see Table 5). Because the measures were those of change, the student characteristics in this

TABLE 3
LINEAR REGRESSIONS: EVALUATION (SET) BY RESEARCH VARIABLES

	B	Beta	t	Sig	Tolerance
GPP, Week 16	0.72	.437	11.01	.000	0.38
Like, Week 16	0.48	.385	9.97	.000	0.41
Expected grade	0.36	.116	3.22	.001	0.47
Halo effect	0.52	.080	3.01	.003	0.86
Age	0.02	.026	0.99	.321	0.89
GPA	0.07	.017	0.51	.610	0.53
Sex	0.04	.012	0.47	.641	0.90
Sophomore	−0.02	−.004	−0.16	.877	0.87
Heard	−0.02	−.005	−0.21	.837	0.96
Hard class	0.30	.077	2.80	.005	0.80
Hard grader	0.24	.060	2.23	.026	0.84
Initial expectations	−0.05	−.015	−0.53	.598	0.78
Final grade	−0.07	−.029	−0.77	.441	0.44
Constant	−2.38		−3.46	.001	

NOTE: SET = student evaluation of teaching; GPP = global personality positive; GPA = grade point average.
$r^2 = .757$, $F(13,402) = 96.27$, $p < .0001$

FIGURE 1:   Percent of Variance (r^2) of Final Evaluation of Global Personality Positive (GPP) during Term

TABLE 4
CORRELATION MATRIX OF THE FINAL EVALUATION WITH GLOBAL PERSONALITY POSITIVE (GPP) BY WEEK OF TERM

Variable	Eval	Week 0	Week 1	Week 10	Week 16
1. Evaluation	—	.148*	.323*	.602*	.803*
2. GPP Week 0		—	.479*	.323*	.246*
3. GPP Week 1			—	.499*	.444*
4. GPP Week 10				—	.714*
5. GPP Week 16					—

* Significance at $p < .01$.

analysis were redundant but were included as a type of validity check. A regression of the five personality dimensions shows that the change in each is a significant predictor of the change in the evaluations during the same time period (see Table 6).

It is possible that whatever association may exist between changes in perceived personality and changes in the evaluations may be because of a class effect. The change in GPP between the 10th and 16th weeks was recoded as *down, no change,* or *up.* Each class had some change during this time period in all three categories. An ANOVA was run using class-by-grade change (a 14 × 3 factor analysis: class by direction of personality change). Both personality change and the class effect were significant, but there was no interaction. Personality change accounted for over six times more variance than what was contributed by class differences.

4) How much of the change in personality, resulting in a change in the evaluations, is related to factors such as general likeability and halo effects? There is no previous
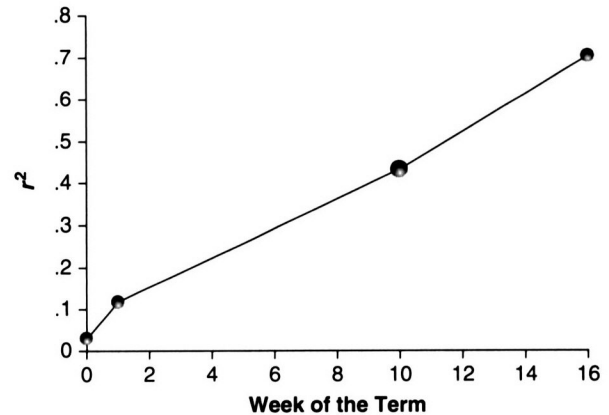
TABLE 5
LINEAR REGRESSIONS: CHANGE IN EVALUATIONS (LAST 6 WEEKS) BY RESEARCH VARIABLES

	B	Beta	t	Sig	Tolerance
GPP change	0.57	.391	8.77	.000	0.72
Likeability change	0.20	.230	5.30	.000	0.76
Expected grade change	0.52	.263	6.46	.000	0.86
Halo change	0.10	.028	0.73	.468	0.93
Age	0.03	.044	1.09	.276	0.90
GPA	−0.12	−.048	−0.92	.357	0.53
Sex	−0.01	−.004	−0.11	.916	0.91
Sophomore	−0.09	−.028	−0.70	.484	0.88
Heard	−0.16	−.064	−1.64	.101	0.95
Hard class	0.12	.047	1.14	.256	0.83
Hard grader	0.10	.011	1.00	.316	0.87
Initial expectations	0.14	.068	1.60	.110	0.79
Final grade	−0.03	−.020	−0.42	.675	0.61
Constant	−0.85		−1.37	.172	

NOTE: GPP = global personality positive; GPA = grade point average.
$r^2 = .442$, $F(13,392) = 23.90$, $p < .0001$

research to guide us here, so what follows is exploratory in nature. Mathematically, the measures at Week 16 are equal to those at Week 10 plus the change from Week 10 to Week 16. A logical starting place to investigate the personality change, then, would be the conditions that existed at Week 10. The initial regression shown in Table 7 tests Clayson and Haley's (1990) contention that likeability is the essential scale in SET along with Orsini's (1988) contribution of a halo effect as contributors to GPP change. Both were found to be positive predictors of the personality change, along with GPP at Week 10. The expected grade at Week 10 was not related to personality change.

TABLE 6
LINEAR REGRESSIONS: CHANGE IN EVALUATIONS (LAST 6 WEEKS) BY CHANGE IN PERSONALITY VARIABLES

	B	Beta	t	Sig	Tolerance
Agreeableness	0.27	.277	6.12	.000	0.73
Creativity	0.20	.213	5.04	.000	0.84
Conscientiousness	0.16	.166	3.86	.000	0.81
Stability	0.10	.119	2.84	.005	0.84
Extroversion	0.09	.088	2.20	.028	0.94
Constant	−0.22		−4.86	.000	

NOTE: Variables arranged by the magnitude of beta.
$r^2 = .321$, $F(5,456) = 43.07$, $p < .0001$

The second regression in Table 7 adds the student characteristics. None of these significantly affect the measure of GPP change. The third regression looks at the existing conditions at Week 10 and then adds the change that occurred in likeability, expected grade, and halo effects. The fourth regression looks at only the variables that remained significant in the previous analyses. More than 48% of the variance of the change in GPP can be accounted for by preexisting GPP and likeability factors plus change in likeability, expected grades, and halo.

How does this change in GPP relate to the final evaluation? A path analysis was run using LISREL, assuming that expected grade and halo effects at Week 16 and preexisting GPP and GPP change were related directly to the final evaluations (see Figure 2). All the paths in the model have coefficients significant at less than the .01 level. The standardized coefficient for the path between GPP Difference and Evaluation is .63 ($t = 8.62$). The fit of the data is very good (AGFI = 0.94), with the path analysis accounting for 73% of the variance of the final evaluation.

Assuming that change must begin from some foundational point, here defined as the GPP measure at Week 10, the change in GPP measures can be seen as resulting from changes in the students' perception of expected grade, halo effects, and whatever modifies the general level of liking for the instructor.

DISCUSSION

The study attempted to answer four questions.

1) Does a relationship exist between personality characteristics and evaluation in marketing and business core classes? The study found a consistent and positive relationship between course and instructor evaluations and personality measures. The effect was not significantly modified by any student variable tested, including GPA, gender, age, initial class expectations, final class performance, or the students' perception of the difficulty of the class or the rigor of the instructor's grading standards. This implies that students universally are associating perceived personality with instructional effectiveness.

2) If a relationship does exist between personality and evaluation, how early in the term does it develop? Within

TABLE 7
LINEAR REGRESSIONS: PREDICTING THE CHANGE IN GLOBAL PERSONALITY POSITIVE (GGP WEEK 16 – GGP WEEK 10)

	B	Beta	t	Sig	Tolerance
Initial conditions					
GPP Week 10	−.35	−.425	−6.66	.000	0.49
Likeability Week 1	.16	.268	4.30	.000	0.51
Expected grade Week 10	−.05	−.036	−0.70	.486	0.93
Halo Week 10	.10	.032	0.70	.486	0.93
Constant	1.15		4.13	.000	
$r^2 = .093$, $F(4,458) = 11.73$, $p < .0001$					
Initial conditions plus student characteristics					
GPP Week 10	−.32	−.412	−6.02	.000	0.48
Likeability Week 10	.15	.262	3.93	.000	0.50
Expected grade Week 10	−.15	−.116	−1.96	.051	0.63
Halo Week 10	.09	.032	0.64	.523	0.90
Age	.01	.035	0.70	.485	0.90
GPA	.10	.060	0.94	.349	0.54
Sex	−.04	−.026	−0.52	.604	0.92
Sophomore	−0.12	−.054	−1.58	.116	0.89
Heard	−0.17	−.1.03	−2.14	.033	0.96
Initial expectations	−0.01	−.009	−0.17	.864	0.76
Final grade	0.12	.108	1.64	.102	0.52
Constant	0.61		1.13	.260	
$r^2 = .115$, $F(11,395) = 64.67$, $p < .0001$					
Initial conditions plus change					
GPP Week 10	−.53	−.645	−13.07	.000	0.46
Likeability Week 10	.39	.665	12.52	.000	0.40
Expected grade Week 10	.04	.033	0.92	.356	0.87
Halo Week 10	.28	.092	2.15	.032	0.62
Likeability difference[a]	.37	.633	15.97	.000	0.71
Halo difference	.44	.174	4.18	.000	0.65
Expected grade difference	.16	.116	4.18	.002	0.84
Constant	0.62		2.84	.004	
$r^2 = .490$, $F(7,455) = 62.56$, $p < .0001$					
Combination of significant predictors[b]					
Likeability Week 10	.39	.673	12.67	.000	0.40
Likeability difference	.38	.649	16.55	.000	0.73
GPP3	−.51	−.620	−12.83	.000	0.48
Halo difference	.31	.121	3.55	.000	0.97
Expected grade difference	.15	.107	3.09	.002	0.94
Constant	0.53		3.38	.000	
$r^2 = .484$, $F(5,457) = 85.88$, $p < .001$					

a. Difference score from Week 16 minus Week 10.
b. Variables rearranged by magnitude of beta coefficient.

fewer than 5 minutes of initial contact, students' perception of the personality of the instructor is associated with the final evaluation given 16 weeks later. The students had not yet seen the course syllabus or been exposed to any pedagogical interchange. The association grew stronger over the
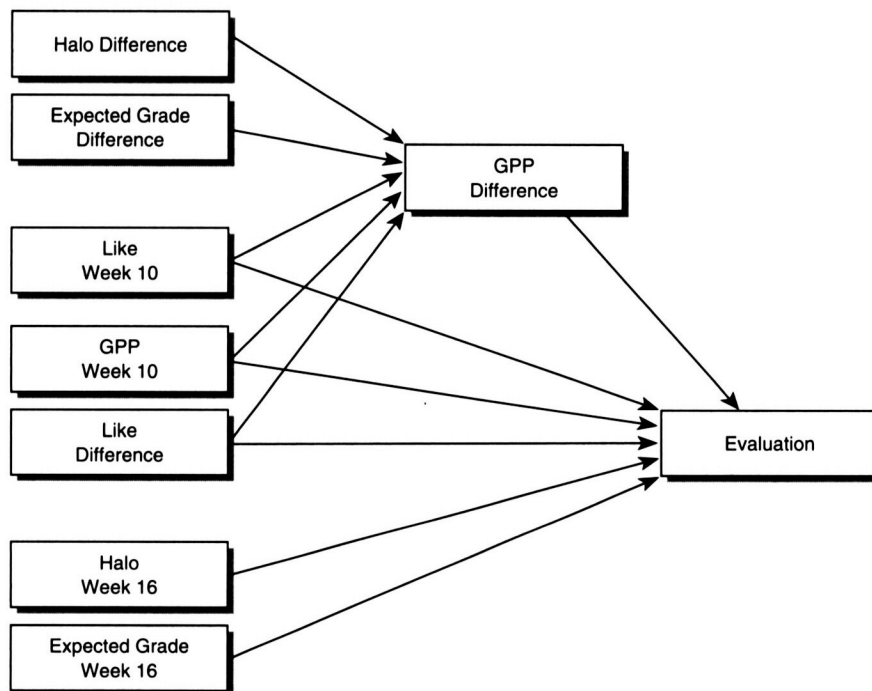
FIGURE 2:   Path Diagram
NOTE: GPP = global personality positive.

term until the global personality perception accounted for fully 64% to 73% of the variance of the evaluation.

3) Will changes in personality after a class is well established be related to changes in the evaluations given? When the students were fully familiar with the instructor, the classroom environment, and their own grades after 10 weeks of class, subsequent changes in the perception of personality were associated significantly with changes in the evaluations. The association appears to be robust and cannot be attributed logically to any persistent classroom variable. Personality, by definition, however, should not be changing rapidly over time, especially in a mere 6 weeks. Nevertheless, almost 60% of the students reported changes in their perception of the instructor's personality that were greater than what would be expected by mere rounding errors.

4) How much of the change in personality, resulting in a change in the evaluations, is related to other factors, such as general likeability and halo effect? The analysis showed that much of the change in personality can be accounted for by changes in expected grades and halo effects combined with a measure of likeability. How much of the measures of personality perception before Week 10 could be attributed to the same factors was not determined by this study.

Limitations

The data were gathered at one business school. Other settings may show more or less of the effects shown here.

However, the classes chosen in the study were business core classes consisting of students from eight different majors. In addition, the percent of variance in SET accounted for by personality was firmly within the range established by Murray (1975), Feldman (1986), Sherman and Blackburn (1975), and Clayson and Haley (1990).

All of the Big Five dimensions were found to be significantly related to SET except extroversion. This trait generally has been found to be associated with the evaluations (Erdle, Murray, and Rushton 1985). In isolation, extroversion was highly related to the total evaluation, but in a regression with other personality dimensions, it was reduced to nonsignificance. Perceived change in extroversion was found, however, to be associated with the change in the evaluations. Extroversion did not have less variation than other personality factors, nor was it more or less associated with the halo effect. In fact, the existence of a nonsignificant personality factor logically removes the halo effect as the only influence of the relationship between personality and evaluation.

Implications

The underlying dynamics of the high degree of association between measures of personality and SET can be seen in a variety of ways. Finding an association between the final evaluation of the class at the end of the term and a personality evaluation made within 5 minutes of exposure, as well as corresponding changes in personality and evaluations in the last weeks of the term, makes a validity argument for

the relationship between personality and evaluation difficult to defend. Sherman and Blackburn (1975) suggested that personal characteristics are the cause of the perceived instructional effectiveness. Erdle, Murray, and Rushton (1985) claimed that instructors' personalities are reflected in certain classroom teaching behaviors that in turn are validly rated by students. The findings of this study make both claims untenable. Feldman (1986) suggested that a relationship between personality and evaluation might exist if both were related to other variables that normally may or may not be considered to be indicators of instructional effectiveness. Hence, both personality and the evaluations would be contaminated. This suggestion was not contradicted by the findings of the study. Personality changes associated with evaluation changes not only were observed in the last part of a term, but different students perceived changes going in opposite directions in the same instructor during the same time period.

The findings suggest that a fourth explanation, one not hypothesized by previous researchers, should be considered. It is possible that personality, as perceived by the students, and whatever it is that SET is measuring may be essentially the same construct. Arguments about the validity of personality as a measure of some independent SET, contamination hypotheses, or discussions of what causes one or the other simply may be misplaced. Combining the expected grade and a halo effect to personality and likeability accounted for an astonishing 73% of the total variance of SET, an association rarely found in any measurement of human behavior or perception. It appears that questions asked of the students pertaining to how they would evaluate their instructor's effectiveness, including questions about their perception of the learning environment, grading standards, and satisfaction with learning, could be replaced with a personality inventory of the instructor with little change in outcome. As a preliminary test of this hypothesis, a confirmatory factor analysis was run with the present data. A canonical correlation of 0.92 between a factor consisting of the Big Five personality constructs and likeability and a factor composed of the individual measures of evaluation was found, thus reinforcing the contention made by Clayson and Haley (1990) that SET scales could be seen as a single measure of something they called *likeability*. Only further research could confirm this hypothesis.

The results of this study also help to resolve a paradox found in previous research. Langbein (1994) used an established SET form to measure the effects of a number of variables on the SET measure. The SET form could be reduced to one factor and the items could be summed with a Cronbach's alpha of 0.99. The author claims this is strong evidence that random measurement error is not the source of the SET variance. Yet the variables in the study accounted for only about 12% of the variance of SET. It appears that the addition of the students' perception of the instructor's personality accounts for the major share of Langbein's missing variation.

It is not known if the personality-evaluation effect is related to a projection of the students' own traits on their instructors. Students have been found to project gender biases on their teachers (Basow 2000; Langbein 1994). In addition, the students' own locus of control affects the evaluations (Grimes, Millea, and Woodruff 2004) along with preferences for conforming social styles (Schlee 2005), but the students' personality has not been found to be directly associated with SET (Marsh and Roche 1997). Further, matching students' personality types with the instructor's does not appear to enhance performance in business students (Ziegert 2000).

SET Validity

The finding that SET is largely a measure of student-perceived personality does not in itself invalidate the instruments. As Feldman (1986) pointed out, a certain amount of validity overlay could be expected. We could assume that an environment that the students perceive as accepting, warm, and energetic would be conducive to learning. In fact, interviews of marketing students indicate that their first priority when looking at instruction is that instructors are empathetic, caring, and understanding, thus "sustaining the human interface" (Faranda and Clark 2004, p. 280). The problem, however, is highlighted by an example given by Simpson and Siguaw (2000). They wrote of a professor at a small university who managed to raise students' scores on a nationally scaled exam in finance from the 13th to the 97th percentile, but the instructor consistently placed in the lowest third of all faculty on the SET. If good instruction is defined by a standardized learning measure, then this instructor is an outstanding teacher. On the other hand, if good teaching is simply what a student evaluation says it is, then this professor probably should be replaced. As pointed out by Clayson and Haley (2005), his or her replacement surely will not commit the same error. This is, in summary, the conundrum of perceived personality accounting for so much of the variance of SET. The relationship, in and of itself, does not invalidate the instruments as far as this perception reflects important components of instruction. It does, however, have the potential of invalidating the evaluations if the evaluations are used blindly and singularly to measure the quality of instruction. Chonko (2004) highlighted the problem in a discussion of quackery in marketing education. A quack instructor will attempt to manipulate the evaluations by using tactics that have little to do with learning or achievement. Consequently, the evaluations "appear to be a false metric."

On an even darker note, if we make the assumption that student perceptions are even marginally related to relatively long-lasting traits, it may be true that some teachers never will receive consistently high evaluations in certain environments, irrespective of anything they do or possibly could do. To the extent that instruction is judged by personality and the perceptions reflect reality, teaching may be seen as a vocation to which certain individuals are simply born.

Attempts in graduate schools to teach new PhDs to be adequate teachers would be justified, but attempts to produce master teachers would be a waste of precious time.

Machina (1987) addressed the personality issue 20 years ago. He did not deny that personality factors might influence the evaluations: "Consequently, some instructors will not receive the high marks they deserve from their students for careful course planning, conscientious review of student work, and brilliantly conceived lectures" (p. 22). Machina's reaction to this was hardheaded. Because he was a strong advocate of SETs and did not want them eliminated, he stated, "It is regrettable if some fine scholars are unable to interact with certain groups of students because of personality factors, but however unfair, such a fact is nevertheless educationally relevant" (p. 21). We do not agree. While it may be relevant as long as student evaluations are used to evaluate teaching, it need not be the case if other measures of teaching effectiveness, such as standardized learning outcomes, are used. Student evaluations of instruction appear to follow a seriously flawed paradigm. At the very least, they should be closely monitored both by faculty and by administrators when they are used as indicators of teaching quality. In the long term, research and discussions should be initiated to replace the current SET system with some other form of evaluation. A consensus needs to be reached about what good teaching and instruction are, and methods should be derived to scale that concept. The new system need not exclude student perceptions. Whether an instructor is liked is an important piece of information, but using only this information reduces the evaluation process to a beauty contest in which the students are ultimately the biggest losers.

REFERENCES

Allport, Gordon W. 1955. *Becoming: Basic considerations for a psychology of personality (Terry lectures)*. New Haven, CT: Yale University Press.

Ambady, Nalini, and Robert Rosenthal. 1993. Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology* 64 (3): 431–41.

Basow, Susan A. 2000. Best and worst professors: Gender patterns in students' choices. *Sex Roles: A Journal of Research* (Sept.) Retrieved March 31, 2006, from www.findarticles.com/p/articles/mi_m2294/is_2000_sept/ai_71118815.

Boice, Robert 1992. Countering common misbeliefs about student evaluation of teaching. *ADE Bulletin* 101 (Spring): 2–8.

Braskamp, Larry A., and John C. Ory. 1994. *Assessing faculty work: Enhancing individual and institutional performances*. San Francisco, CA: Jossey-Bass.

Buchanan, Tom 2001. Online implementation of an IPIP Five Factor personality inventory. Retrieved from //users.wmin.ac.uk/~buchant/wwwffii/introduction.html (Last accessed April 6, 2005).

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.

Centra, John A. 1993. *Reflective faculty evaluations: Enhancing teaching and determining faculty effectiveness*. San Francisco, CA: Jossey-Bass.

Chonko, Lawrence B. 2004. If it walks like a duck . . . : Concerns about quackery in marketing education. *Journal of Marketing Education* 26 (1): 4–16.

Clayson, Dennis E. 1989. Halo effects in student evaluations of faculty: A question of validity. In *Marketing: Positioning for the 1990s, Proceedings of the annual meeting of the Southern Marketing Association*, edited by Robert L. King, 271–75. Charleston, SC: Southern Marketing Association and The Citadel.

———. 1999. Students' evaluation of teaching effectiveness: Some implication of stability. *Journal of Marketing Education* 21 (1): 69–75.

———. 2004. A test of reciprocity effects in the student evaluation of instructors in marketing classes. *Marketing Education Review* 14 (2): 11–21.

Clayson, Dennis E., Taggart F. Frost, and Mary Jane Sheffet. 2006. Grades and the student evaluation of instruction: A test of the reciprocity effect. *Academy of Management Learning & Education* 5 (1): 52–65.

Clayson, Dennis E., and Debra A. Haley. 1990. Student evaluations in marketing: What is actually being measured? *Journal of Marketing Education* 12 (3): 9–17.

———. 2005. Marketing models in education: Students as customers, products, or partners. *Marketing Education Review* 15 (1): 1–17.

Digman, John M. 1990. Personality structure: An emergence of the five-factor model. *The Annual Review of Psychology* 41: 417–40.

Erdle, Stephen, Harry G. Murray, and J. Philippe Rushton. 1985. Personality, classroom behavior and student ratings of college teaching effectiveness: A path analysis. *Journal of Educational Psychology* 77 (4): 394–407.

Faranda, William T., and Irvine Clark III. 2004. Student observations of outstanding teaching: Implications for marketing education. *Journal of Marketing Education* 26 (3): 271–81.

Felder, Richard M. 1995. What do they know anyway? *Chemical Engineering Education* 26 (3): 134–35.

Feldman, Kenneth A. 1986. The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: A review and synthesis. *Research in Higher Education* 24 (2): 139–213.

Foote, David A., Susan K. Harmon, and Donna T. Mayo. 2003. The impacts of instructional style and gender role attitude on students' evaluation of faculty. *Marketing Education Review* 13 (Summer): 9–19.

Grimes, Paul W., Mehan J. Millea, and Thomas W. Woodruff. 2004. Grades—Who's to blame? Student evaluations of teaching and locus of control. *Journal of Economic Education* 35 (2): 129–47.

Hewett, L. J., Garvin Chastain, and Steven Thurber. 1988. Course evaluations: Are students' ratings dictated by first impressions? Paper presented at the Rocky Mountain Psychological Association, April 21–22, Snowbird, Utah.

Jang, Kerry L., Robert R. McCrae, Alois Angleitner, Ranier Riemann, and W. John Livesley. 1998. Heritability of facet-level traits in a cross-cultural twin sample: Support for a hierarchical model of personality. *Journal of Personality and Social Psychology* 74 (6): 1556–65.

Johnson, Valen E. 2003. *Grade inflation: A crisis in college education*. New York: Springer.

Kulik, James A. 2001. Student ratings: Validity, utility, and controversy. *New Directions for Instructional Research* 109 (Spring): 9–25.

Langbein, Laura I. 1994. The validity of student evaluations of teaching. *PS: Political Science and Politics* 27 (3): 545–52.

Lantos, Geoffrey P. 1997. Motivating students: The attitude of the professor. *Marketing Education Review* 7 (2): 27–38.

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): 108–19.

Marsh, Herbert W., and Dennis Hocevar. 1991. Students' evaluation of teaching effectiveness: The stability of mean rating of the same teachers over a 13-year period. *Teaching & Teaching Education* 7 (1): 303–14.

Marsh, Herbert W., and Lawrence A. Roche. 1997. Making students' evaluations of teaching effectiveness effective. *American Psychologist* 52 (11): 1187–97.

———. 1999. Reply upon SET research. *American Psychologist* 54 (7): 517–18.

———. 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): 202–28.

McCrae, Robert R., and Paul T. Costa Jr. 1997. Personality trait structure as a human universal. *American Psychologist* 52 (5): 509–16.

Murray, Harry G. 1975. Predicting student ratings of college teaching from peer ratings of personality types. *Teaching of Psychology* 2: 66–70.

Murray, Harry G., J. Philippe Rushton, and Sampo V. Paunonen. 1990. Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology* 82 (2): 250–61.

Naftulin, Donald H., John E. Ware, and Frank A. Donnelly. 1973. The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education* 48 (July): 630–35.

Orsini, Joseph L. 1988. Halo effects in student evaluations of faculty: A case application. Journal of Marketing Education 10 (2): 38-45.

Ortinau, David J., and Robert P. Bush. 1987. The propensity of college students to modify course expectations and its impact on course performance information. *Journal of Marketing Education* 9 (Spring): 42–52.

Osgood, Charles E., George J. Suci, and Percy H. Tannenbaum. 1957. *The measurment of meaning.* Urbana, IL: University of Illinois Press.

Sauber, Matthew H., and R. Rodman Ludlow. 1988. Student evaluation stability in marketing: The importance of early class meetings. *The Journal of Midwest Marketing* 3 (1): 41–49.

Schlee, Regina P. 2005. Social styles of students and professors: Do students' social styles influence their preferences for professors? *Journal of Marketing Education* 27 (2): 130–42.

Sherman, Barbara R., and Robert T. Blackburn. 1975. Personal characteristics and teaching of college faculty. *Journal of Educational Psychology* 67 (1): 124–31.

Simpson, Penny M., and Judy A. Siguaw. 2000. Student evaluations of teaching: An exploratory study of the faculty response. *Journal of Marketing Education* 22 (3): 199–213.

Soldz, Stephen, and George E. Vaillant. 1999. The Big Five personality traits and the life course: A 45-year longitudinal study. *Journal of Research in Personality* 33 (2): 208–32.

Stumpf, Stephen A., and Richard D. Freedman. 1979. Expected grade covariation with student ratings of instruction: Individual versus class effects. *Journal of Educational Psychology* 71 (3): 293–302.

Theall, Michael, and Jennifer Franklin. 2001. Looking for bias in all the wrong places: A search for truth of a witch hunt in student ratings of instruction? *New Directions for Institution Research* 109 (Spring): 45–56.

Widmeyer, W. Neil, and John M. Loy. 1988. When you're hot, you're hot! Warm-cold effects in first impressions of persons and teaching effectiveness. *Journal of Educational Psychology* 80 (1): 118–21.

Ziegart, Andrea L. 2000. The role of personality temperament and student learning in principles of economics: Further evidence. *Research in Economic Education* 31 (4): 307–22.