The Relative Influence of Published Teaching Evaluations and Other Instructor Attributes on Course Choice

Wendy Bryce Wilhelm

One hundred and twenty-seven undergraduate business majors at a regional northwestern university completed a series of hypothetical choice tasks in which they were asked to choose between two courses that varied with respect to the instructor's course evaluations, grading leniency, the course's worth/usefulness, and the assigned workload. Data analysis revealed that while evaluations exert a significant influence on choice, course worth and grading leniency are the most important determinants of course choice. Share of preference simulations indicated that students are twice as likely to choose a course with an instructor who receives excellent, as opposed to average, course evaluations, all else being equal. However, students are willing to put up with poor course evaluations or a heavy workload if they believe that they will gain a great deal of useful knowledge. The article concludes with a call for more research on the decision process relating to course choice.

Keywords: *course selection decisions; student evaluations of teaching; conjoint experiment; teaching effectiveness; market simulation*

Research on the validity and utility of student evaluations of teaching (SET) has been ongoing since they were first administered in 1926 at the University of Washington (d'Apollonia and Abrami 1997). The question of whether SET are a valid measure of teaching effectiveness and issues relating to their use in faculty evaluation and promotion decisions has generated well over 2,000 separate pieces of academic research and commentary (Wilson 1998).

While research on the validity and use of SET by administrators and faculty is plentiful, little research has been done on how students perceive and use SET. One study found that students believe SET are important and should affect faculty advancement decisions (Ahmadi, Helms, and Raiszadeh 2001). Some call for more research on how students view the process of completing course evaluations and note that "students must accept the necessity of acting responsibly in rating their instructors" (Ory and Ryan 2001, p. 41).

However, few studies have addressed the issue of how students might use SET to select courses, even though one of the stated objectives of gathering evaluation data is to improve students' course selection decisions (Marsh and Roche 1997; McKeachie 1997). A growing number of universities publish SET online, and anecdotal data suggest that student demand for, and usage of, this easily accessible information is high (Haskell 1997c; Nana Lowell, e-mail message to author, February 26, 2003; Tarleton 2003). Recent attempts to deny student access to SET by faculty and administrators at the University of Wisconsin and the University of Idaho have failed in the courtroom (Haskell 1997c, note 55), although some question the legality of releasing SET to the public on the grounds that the publication of student "anecdotal data" may libel or defame a faculty member (Haskell 1997v).[1] In sum, it appears that an increasing number of students on U.S. campuses are, or soon will be, able to easily access past teaching evaluations for courses in which they are considering enrolling, despite attempts by some groups to restrict availability.

The decision process that students engage in when selecting a course is frequently a complex one characterized by high involvement and perceived risk, particularly when the decision context requires a choice among courses or course sections within a major field of study (Babad, Darley, and Kaplowitz 1999). Research directed at increasing our understanding of this process may contribute to the development of tools that can improve course selection decisions (e.g., computer-assisted course selection software). Such improvements in course choice may, in turn, increase students' satisfaction levels with the college experience and influence the

Wendy Bryce Wilhelm is an associate professor of marketing in the College of Business and Economics at Western Washington University, Bellingham.

Journal of Marketing Education, Vol. 26 No. 1, April 2004 17-30
DOI: 10.1177/0273475303258276

range of occupational opportunities available to them upon graduation.

The conjoint study discussed in this article explores how business students might use SET and other beliefs or perceptions about instructors to make course-selection decisions in their major. Students were exposed to a series of hypothetical choice tasks where they were asked to choose between two sections of a course that varied with respect to the instructor's SET ratings, the instructor's perceived grading leniency, perceived course worth or the usefulness of the knowledge provided by the instructor, and the perceptions regarding assigned workload.

LITERATURE REVIEW AND RESEARCH QUESTIONS

Validity and Reliability of SET

To date, there has been little agreement among researchers about how to measure teaching effectiveness or on whether current measures accurately reflect an instructor's ability to teach (Greenwald 1997; Theall and Franklin 2001). Some suggest that SET may measure student satisfaction (Cahn 1987) or instructor popularity (Wilson 1998), rather than an instructor's ability to help students learn. There appears to be some agreement that student rating forms should be multidimensional, and studies using structural equation modeling and/or factor analysis have identified up to 28 dimensions of teaching effectiveness (Feldman 1989; Marks 2000; Marsh and Bailey 1993). These dimensions include instructor liking/concern, perceived learning, workload/difficulty, expected grade, course organization and planning, and course worth (i.e., utility of knowledge gained, relevance to major).

Other researchers have focused on issues of measurement reliability, reporting that many extraneous factors can influence or bias SET. These include the time of day evaluations are administered (Nichols and Soper 1972), class size and status (required versus elective; Braskamp and Ory 1994; McKeachie 1997), pedagogical style (Davis, Shekhar, and Van Auken 2000), the instructor's grading leniency (Bacon and Novotny 2002; Greenwald and Gillmore 1997), and the sex and rank of the instructor (Freeman 1994; Marsh and Hocevar 1991). There is considerable overlap between the "dimensions" identified above and these "influences," making it difficult to distinguish between them (see Marks [2000] and Ory and Ryan [2001] for a discussion of this point). Table 1 summarizes the research findings on key factors that have been found to influence student course evaluations.

Factors Influencing Course Choice

The trend toward giving students greater access to instructors' SET increases the need for more research on the validity and reliability issues raised above. Of equal importance is the need for more research on how students would or do use these published evaluations to choose courses and the relative importance of SET compared to other factors. In this study, interest centers on how instructor factors, like SET or perceived grading leniency, affect course choice. Thus, the choice situation of most interest here is when the student has the option of selecting among several sections of a required course in his or her major that are taught by different instructors, thereby controlling for differences in subject matter and personal interest (e.g., a required marketing research course versus an elective advertising course). Other possible influence factors related to course choice, such as time of day, were held constant in this study.

Student evaluations of teaching (SET). With the exception of one recent study (Babad, Darley, and Kaplowitz 1999), studies on the relationship between SET and course selection among similar courses or section selection in multisection courses (most relevant for the present study on instructor influences) were conducted in the 1970s and 1980s (Coleman and McKeachie 1981; Leventhal, Abrami, and Perry 1976; Leventhal et al. 1975; Martin 1989). Findings from this earlier program of research indicate that (1) SET are primarily a measure of instructor liking and reputation (wit, enthusiasm, personality, expertise, approachability); (2) sections with higher SET ratings are more frequently selected, even if the workload in those sections is believed to be heavier; and (3) instructor reputation is more important in section selection than other course factors, especially for those students with lower educational aspirations. Most of these studies were with freshmen, precluding an assessment of how the relative importance of SET in course choice might change as students progress through college.

A more recent study by Babad, Darley, and Kaplowitz (1999) corroborated these earlier findings. The authors also found that instructor reputation and liking become *less* important in course selection decisions as students progress from 100- to 400-level courses, while perceived course worth (the usefulness of the knowledge gained from an instructor) becomes more important, an interesting developmental trend. Taken together, SET as a measure of instructor popularity or reputation appear to have a significant influence on course selection, but their relative importance may decrease as students begin taking advanced courses in their major.

> *Research Question 1:* Do published student evaluations for courses have a significant influence on business students' course choices when they must choose between sections of a required course in their major that are taught by different instructors?

Perceived course workload and grading leniency. It is reasonable to expect that those factors or attributes that influence SET also influence students' course choices. Course workload and grading leniency can have a significant positive or negative influence on SET (see Table 1), although evidence

TABLE 1
SUMMARY OF RESEARCH ON FACTORS THAT INFLUENCE STUDENT EVALUATIONS OF TEACHING (SET)

Factor	*Influence (size/direction) on SET*	*References*
Instructor liking/enthusiasm	Strong positive influence	Marks (2000); Radmacher and Martin (2001); Williams and Ceci (1997); Wilson (1998)
Course workload/difficulty	Strong positive or negative influence; workload inversely related to grading leniency and expected grade	Bacon and Novotny (2002); D'Apollonia and Abrami (1997); Greenwald and Gillmore (1997); Marks (2000); Marsh and Roche (1997)
Grading leniency	Strong positive influence (but may be moderated by other factors such as workload, undergrade/graduate status, and motivation level)	Bacon and Novotny (2002); Greenwald and Gillmore (1997); Marsh and Roche (1997); McKeachie (1997)
Course worth/relevance	Strong positive influence	Bacon and Novotny (2002); Feldman (1989); Marsh and Bailey (1993)
Expected grade	Strong positive influence	Archibold (1998); Barnoski and Sockloff (1976); Greenwald and Gillmore (1997)
Pedagogical style	Mixed findings but may have moderate influence on SET	Davis, Shekhar, and Van Auken (2000)
Sex of instructor	Weak influence (female students rated lower than male students)	Feldman (1992); Freeman (1994); Theall and Franklin (2001)
Rank and experience of instructor	Moderate positive or negative influence	Clayson (1999); Marsh and Hocevar (1991); Pohlmann (1975)
Class size	Weak negative influence	Avi-Itzhak (1982); Braskamp and Ory (1994); McKeachie (1997)
Course status: required versus elective	Elective courses rated more favorably	Costin, Greenough, and Menges (1971); Feldman (1978); McKeachie (1997)
Time of day course meets (p.m. ratings lower than a.m. ratings)	Weak influence	Nichols and Soper (1972)

for the size and direction of their effects on course selection is inconclusive (Babad, Darley, and Kaplowitz 1999; Coleman and McKeachie 1981; Takeshita and Maeda 1999). For example, Takeshita and Maeda (1999), in an empirical study designed to evaluate a university Web-based course selection and registration system, found that when it came to the selection of courses, students chose those courses "whose instructors are popular for one reason or another and whose credits are easier to obtain" (p. 997), suggesting that workload is negatively related to course choice. On the other hand, Babad, Darley, and Kaplowitz (1999) found that grading leniency and workload exerted relatively little influence on course selection regardless of course level (e.g., 100, 400). Furthermore, perceived workload and grading leniency were positively related to course choice for some students and courses, and negatively associated with course choice in other cases. These two instructor attributes are included in the present study due to their strong influence on SET, the current controversy surrounding the direction of that influence, and the inconclusive evidence concerning their effects on course choice.

Perceived course worth. This instructor attribute is defined as the perceived usefulness or relevance of the knowledge gained in a course from a particular instructor. For business students, course worth typically refers to the "real-world" orientation of the instructor's lectures and assignments, and many students perceive "theory as the antithesis of reality" (Bacon and Novotny 2002, p. 7). Research with business students suggests that perceived course worth or relevance has an important influence on SET and students' satisfaction with a course, and that worth is a function of instructor-controlled course components such as the type of assignments, number of guest lecturers by practitioners, and the amount of lecture material focusing on real-world examples and applications (Bacon and Novotny 2002; Karns 1993; Kelley, Conant, and Smart 1991). This emphasis on relevance is understandable, since many students begin the 2-year business program in their junior year, at a time when they are becoming more serious about the courses they take and how well a course prepares them for the career they envision. Babad, Darley, and Kaplowitz (1999) reported a similar developmental trend across all majors, that is, course worth is a more important

factor in course selection for 300- and 400-level courses, relative to lower level courses.

Relative influence of the four instructor attributes. This review of previous research supports the thesis that SET, workload, grading leniency, and course worth are key instructor attributes influencing section selection, with each exerting a separate and unique influence on course selection (Babad, Darley, and Kaplowitz 1999; Coleman and McKeachie 1981). Some uncertainty remains, however, regarding the direction and relative influence that each of these factors has on course selection. Since there does seem to be agreement in the literature that SET capture some aspect of instructor liking or popularity, the latter factor is not included as a separate instructor attribute in the present study.

Research Question 2: What is the relative importance of the following factors in choice of a required business course: (1) SET, (2) perceived course workload, (3) perceived grading leniency, and (4) perceived course worth?

Ideal product configuration and trade-offs. The third research question focuses on what the ideal combination of instructor attribute levels looks like to students and what kinds of trade-offs students are willing to make in their selection decisions. Are students willing to accept a heavier workload if they know that the knowledge gained from this instructor will be useful to their major? Are students willing to choose a course with low SET if they know that the instructor is a lenient grader, well known for giving all As and Bs?

Research Question 3: What does the "ideal" course look like to business students, in terms of the preferred level of each attribute? What kinds of trade-offs are students willing to make between levels of each attribute?

METHOD

Research Design

Students' stated preferences for course options were evaluated using conjoint analysis. Conjoint analysis has become one of the most popular multivariate techniques—with both marketing academics and marketing research practitioners—for understanding how consumers develop preferences for products because of its ability to realistically model many choice processes (Carroll and Green 1995; Green and Krieger 2002; Orme 2002). It is based on the premise that consumers evaluate the overall utility of a hypothetical product (e.g., university course) by combining the separate amounts of utility provided by each attribute (e.g., SET, perceived workload). It thus portrays consumers' decisions realistically as trade-offs among multiattribute products (e.g., "I am willing to choose a section/course that receives excellent student ratings, even if I believe the course workload will be heavy").

A questionnaire is used to obtain a respondent's overall evaluations of a set of product concepts that are prespecified in terms of levels of different attributes. External validity is enhanced to the extent that the product attributes reflect important attributes consumers consider in their decision-making process. As a decompositional model, conjoint analysis then "decomposes" the respondent's overall evaluations to uncover the utility value or importance weight he or she places on each attribute and attribute level (Green and Srinivasan 1990). Since the goal of the present study is to understand what attributes influence student preference for hypothetical course "products," conjoint analysis was selected as the most appropriate means of addressing the research questions.

Use of choice-based conjoint (CBC) analysis. A particular type of conjoint analysis, experimental choice or "choice-based conjoint" (CBC) analysis, was developed in the 1980s in response to industry desires to consider explicit competitive contexts (Carroll and Green 1995). More recently, the use of CBC analysis by marketing research practitioners has experienced significant growth (relative to ratings-based conjoint analysis) as "more companies want to understand how people make *choices* [italics added]" (Vence 2003, p. 4). Rather than rate each product concept/profile one at a time on a measure of attractiveness or likelihood of purchase ("ratings-based" conjoint), respondents are asked to choose, that is, make a preference judgment, between a series of two or more competitive product profiles. This approach to measuring preferences combines discrete choice responses, a logit model that is applied to these responses, and a fractional factorial design in order to minimize the number of choices respondents have to make. Unlike more traditional conjoint software, CBC analysis produces aggregate part-worths or utilities for each attribute and level; it does not generate a set of individual utilities for each respondent. This is a shortcoming of the technique if the researcher's goal is to study differences in preference structures across market segments, but it is also an advantage vis-à-vis ratings-based conjoint if examining potential two-way interactions between attributes is of interest.

The popularity of CBC analysis, relative to other ratings-based conjoint approaches, is due to a number of factors: (1) the realism of the choice task for both high- and low-involvement products, that is, consumers make choices among products all the time (Green and Krieger 2002); (2) the fact that interactions among product attributes can be estimated without the necessity of defining the interaction terms a priori (Chrzan and Orme 2000); (3) the development of a strong theoretical foundation for CBC analysis, based on a multinomial logit model of choice (Louviere, Hensher, and Swait 2000;

Louviere and Woodworth 1983); and (4) recent empirical studies that demonstrate the superior predictive accuracy of choice-based analysis relative to ratings- or rankings-based conjoint approaches (Vriens, Oppewal, and Wedel 1998). For these reasons, the present study used Sawtooth Software's CBC System to conduct a full-profile conjoint analysis study (see Carroll and Green [1995] and Deal [2002] for a review of this company's products). A Web-based survey was used to collect the choice data.

Selection of attributes: Pilot study. The selection of the appropriate product attributes to include in the choice task is important to a study's external validity. For that reason, a pilot study with 60 business majors was conducted to confirm the importance of the attributes identified by previous research as being potentially the most important in course choice (Table 1) and to uncover any other attributes that the subject population deemed important. Students were given extra credit to identify key instructor attributes they considered when deciding among sections of a required course in their major (open-end) and to complete a conjoint task with the attributes selected on the basis of prior research. Students also provided feedback on (1) the importance of each of the attributes included in the choice task (1 to 5 scale), (2) the importance of any additional attributes they identified (1 to 5 scale), (3) the ease of understanding the instructions and questions, (4) satisfaction with the visual layout and suggestions for change, and (5) any problems with accessing and moving through the Web questionnaire.

The five conjoint attributes included in the pilot study were published course evaluations, grading leniency, course workload, whether the instructor provides useful knowledge relevant to the student's major (course worth), and instructor sex and rank. The latter attribute included four levels (Male/Female × Lecturer/Tenure-Track Professor) so that the main effects of sex and rank could be isolated. Students do use SET, where available, to evaluate courses, and instructors and respondents in this study were told to assume that published course evaluations for all courses were available on the Web (students are aware that the university is in the midst of implementing this policy). Note that grading leniency, workload, and course worth refer to student *perceptions and beliefs* associated with these attributes, regardless of the source of these beliefs (e.g., word-of-mouth communications, syllabus information). While previous research has found that sex and rank exert a relatively small influence on SET (see Table 1), the sex/rank attribute was included in the present study because informal discussions with business students suggest that sex and rank are important considerations when choosing among business courses. The days and times a course meets are also very important in course choice, but since the focus of the present study is on *instructor* attributes, respondents were asked to assume that the class schedules for all course options presented were equally convenient.

Based on the conjoint results and other findings from the pilot study, modifications were made to the instructions and layout of the survey instrument, and one of the attributes (sex and rank of professor) was dropped from further consideration due to its statistically insignificant effect on course choice. The data revealed no new attributes, and there was a general consensus that the four instructor attributes displayed in Table 2 are the most important ones in choice of a required course section.

Each of the attributes used in the main study had three levels: low, moderate, and high (see Table 2). These levels reflect the differences students perceive to exist among instructors of the same course, based on initial expectations and feedback from the pilot study. The present research site, like many other universities, permits instructor decision-making autonomy regarding section/course structure, grading policy, textbook used, and workload assigned. While the subject matter is similar across sections of a required business course, this autonomy produces a range of attribute levels (low to high) on the attributes of interest in this study. The attribute levels included in Table 2 reflect this reality. The same number of levels was used for all attributes to effect a balanced design (an unequal number of attribute levels can bias estimation of importance weights [Johnson 1996]).

Experimental design and dependent measure. Rather than having each respondent evaluate all possible pairs of product concepts (a practically impossible cognitive task), a fractional factorial, randomized experimental design is typically used to select an optimal set of concepts to present to each respondent. The particular randomized design approach used in the present study is the balanced overlap method. This experimental design employs random sampling with replacement for choosing concepts, permitting some level overlap within the same task (i.e., respondents may have to choose between two courses that have the same workload but differ with respect to grading leniency, etc.). This overlap increases the statistical power of the design/test when testing for attribute interactions by minimizing any potential Type II errors associated with a fractional factorial design (Chrzan and Orme 2000; Vriens, Oppewal, and Wedel 1998). Another one of the strengths of the conjoint software employed, Sawtooth's CBC System, is its ability to develop conjoint questionnaires/designs that are nearly orthogonal, using a randomized design to develop a unique set of questions/concepts for each respondent. Such designs are slightly less efficient than truly orthogonal designs, but they have the offsetting advantage that all two-way interactions between attributes/levels can be measured—an important consideration in the present study.

The experimental design included eight different pairs of product concepts, or eight randomized choice tasks, that were unique to each respondent. Two fixed-choice tasks were also included in the design, that is, the two products presented in each fixed task were the same for all respondents. One of

TABLE 2
ATTRIBUTES AND ATTRIBUTE LEVELS USED IN CONJOINT TASK

Course Evaluation	*Grading Leniency*	*Course Workload*	*Course Worth (utility of knowledge provided by professor)*
Poor	Very easy to get an A or B	Light	Low
Average	Moderately easy/difficult to get an A or B	Moderate	Moderate
Excellent	Very difficult to get an A or B	Heavy	High

these fixed tasks was placed first in the questionnaire and was treated as a "practice" question to allow respondents to gain some familiarity with this type of choice question; respondent data for this first task were not used in data analysis. The remaining fixed-choice task, inserted in the middle of the randomized choice tasks, was used as a holdout task to provide an indication of how well the utility data generated from the randomized tasks predict choices not used in their estimation.

For each choice task, two different product concepts, representing different course options, were presented side by side, and respondents were asked to indicate which one they would choose if they had to register for one of them tomorrow.[2] The actual instructions to the respondents and an example of a choice task can be seen in Figure 1. Within each choice task, the presentation order of the attributes was randomized; in other words, the value taken by the course evaluation attribute was not always presented first, as it is in Figure 1. Following the 10 choice tasks, respondents completed several demographic and attitudinal questions. The survey concluded with an open-ended question soliciting respondents' comments about their personal approach to making course choices.

Ordinary least squares (OLS) were used to test the efficiency of the conjoint experimental design, that is, the precision with which the part-worths for each attribute and level can be estimated. This randomized experimental design had a median statistical efficiency of about 98% relative to a generalized orthogonal design.[3] Furthermore, there was little difference between the standard errors for each main effect (.07 to .08), suggesting that heterogeneity of variance is not a problem.

Sample and Procedure

The target population for this study consisted of business majors at a regional 4-year university in the Northwest (N = 927). Faculty members within the College of Business were asked to give their students extra credit and/or encourage them to participate in the study during the spring quarter of 2002. All faculty members were given the same instruction sheet to read to their students. Once respondents accessed the study (at their convenience, outside of class), the first page was devoted to general instructions explaining the purpose of the study and how the findings might be used by faculty and administrators to develop tools to assist students in their course selection decisions. One hundred and twenty-seven students completed the survey, or 14% of the total business school enrollment that quarter.

RESULTS

Sample Characteristics

The sample consisted of juniors and seniors with a median age of 22 (business majors must be at least 3rd-year students). Sixty percent were female students. Major percentages were as follows: 32% Marketing, 26% Finance, 14% Accounting, 10% Management, 7% Management Information Systems (MIS), 3% Economics, and 10% Other (e.g., no major declared, International Business, Production Management). These percentages approximate the major percentages in the larger business school population, with the exception of the Management and Other categories, which typically represent 18% and 2% of majors, respectively.

Analysis of Conjoint Data: Logit Model

Multinomial logit (MNL) analysis was used to analyze the choice data. Logit was chosen because the form of the dependent and independent variables is categorical and because its structure mimics the nonlinear nature of the impact of marketing effects on choice.[4] Like multiple regression and discriminant analysis, logit seeks "weights" for attribute levels (or for combinations of them, if interactions are included in addition to main effects) that maximize the likelihood of the observed pattern of respondent choices, using probabilities derived from these weights.[5] Those weights are analogous to "preference scores" or "part-worth utilities" in conjoint analysis and are computed so that when the weights corresponding to the attribute levels in each concept are added up, the sums for each concept are related to respondents' choices among concepts (see Ben-Akiva and Lerman 1985; Johnson 1996).

In this study, both main and interaction effects models were developed. Logit analysis of the choice data produced no statistically significant two-way interaction terms. The main effects model was also more robust. The addition of two-way interaction terms did not significantly increase the overall explanatory power of the main effects model, as determined by a chi-square test between the main effects model and several models including some or all of the interaction terms.

If these were the only course section options available for a particular required course in your business concentration, which would you choose? Choose by clicking one of the buttons below.

Each of the two sections offered has the following attributes (assume class size and the day/time each section is offered are the same for both sections):

Section A	*Section B*
Professor and course receive average student ratings, as published on the Web	Professor and course receive excellent student ratings, as published on the Web
Very difficult to get an A or B in this professor's course	Very easy to get an A or B in this professor's course
Light workload assigned by professor	Heavy workload assigned by professor
Professor provides little useful knowledge relevant to my major	Professor provides a great deal of useful knowledge relevant to my major

FIGURE 1: Example of Choice Task

TABLE 3
RELATIVE ATTRIBUTE IMPORTANCES

Attribute	*Relative Importances (%)*	χ	(p *Value)*
Course worth	38	109.3	(<.01)
Grading leniency	31	108.7	(<.01)
Course evaluations	24	39.3	(<.01)
Course workload	7	20.2	(<.01)

NOTE: The relative importance of each attribute was calculated by computing the difference between the largest and smallest part-worth for each attribute, summing the differences and normalizing to 100.

Research Questions 1 and 2: Relative Attribute and Attribute Level Importance

The first research question asks whether published course evaluations have a significant influence on students' selection of a particular section of a required course. The part-worth utilities derived from the logit analysis for each instructor attribute were used to calculate the relative importance of each in course choice. These importance weights and their statistical significance levels are reported in Table 3 and show that course evaluations do have a statistically significant influence on choice.

Research Question 2, concerning the relative importance of each factor or attribute, is also addressed in Table 3. The relative importance of an attribute indicates how much difference a particular attribute can make in the total utility of a "product," such as a course; the difference is the range in the attribute's utility values (see note below Table 3). When choosing between sections of a course, course worth or usefulness has the greatest influence on course choice, with a relative importance of 38%, followed by grading leniency at 31%, and course evaluations at 24%. We can also say that published course evaluation information is approximately two thirds as important in influencing choice as information about course worth (importances are ratio data). Course workload, while statistically significant, is the least important consideration in course selection with a relative importance of 7%.

Research Question 3: Share of Preference for Different Course Configurations

This question asks about the attribute configuration of the "ideal" or most preferred course, and the kinds of trade-offs students would be willing to make between attribute levels. Table 4 contains the average utility values for each attribute level. The most preferred course configuration (the one with the greatest total utility) is one in which the instructor provides a great deal of useful knowledge, is a lenient grader, receives excellent student course evaluations, and assigns a moderate workload.

Trade-offs among attribute levels can be calculated from the average utilities presented in Table 4 (see note below Table 4). A more readily interpretable approach to the question of trade-offs, however, uses these part-worth utilities to simulate specific market conditions in which a given set of course configurations, or "products," are available to choose from. Such simulations produce share-of-preference data for the set of course products specified, where share of preference is defined as what percentage of the respondents would prefer or choose each course, given the set of product courses specified. In this study, Sawtooth Software's Market Simulator, with a randomized first-choice simulation method, was used to obtain share-of-preference data.[6]

Tables 5 and 6 present the shares of preference for various hypothetical course products. All else being equal, students are about twice as likely (58% vs. 32%) to choose a course/section with an instructor who receives excellent, as opposed to average, course evaluations (Table 5). However, students are willing to put up with poor course evaluations or a heavy

TABLE 4
RANKING OF ATTRIBUTE LEVELS BASED ON AVERAGE UTILITY VALUES

	Course Attributes (utilities)			
	Course Worth	*Grading Leniency*	*Course Evaluations*	*Course Workload*
Rank	*(average utility value)*	*(average utility value)*	*(average utility value)*	*(average utility value)*
1	A great deal of useful knowledge (71.83)	Very easy to get an A or B (50.88)	Excellent (43.56)	Moderate (11.01)
2	Some useful knowledge (5.86)	Moderately easy to get an A or B (22.58)	Average (9.91)	Light (7.09)
3	Little useful knowledge (–77.69)	Very difficult to get an A or B (–73.46)	Poor (–53.47)	Heavy (–18.10)

NOTE: Values are arbitrarily scaled to sum to 0 within each attribute, so some utilities must receive a negative value. This does not mean that this level is unattractive; it does mean that attributes with positive utilities are preferred over those with negative utilities. Utilities are interval data; we can say that the increase in preference from an instructor who is a hard grader to one who is an easy grader is *less than* the increase in preference from an instructor/course who provides little useful knowledge to one who provides a great deal. However, we cannot directly compare values *between* attributes to say that two different attribute levels with the same utility value (e.g., moderate workload and average course evaluations) are equally preferred.

workload if they believe that they will gain a great deal of useful knowledge, as indicated by the share-of-preference data in Table 6 for the two course products that provide a "great deal" of knowledge (26% and 20%). In fact, students are 4 times as likely to choose a course with a heavy workload that provides a "great deal " of knowledge (20%) than to choose a course with a light workload that provides "some" knowledge (5%), all else being equal. Another simulation run, comparing instructors who are hard graders with those who are lenient, indicated that students are 10 times more likely to choose a course with a lenient grader, all else being equal.

Other Findings: Differences between Business Majors

The logit and market simulation analyses discussed above were rerun with each major. There were no significant differences in the pattern of findings as reported here, with one exception. For marketing majors (n = 40), course workload, with a relative importance of 2% (compared with 7% for the overall sample), does not have a statistically significant influence on course choice. The small subgroup size of 40 makes any conclusions based on this finding tentative at best, however.

DISCUSSION AND IMPLICATIONS

Student Use of SET in Course Choice

Findings from the present study, consistent with extant research (e.g., Coleman and McKeachie 1981; Leventhal et al. 1975), indicate that published SET information would play a significant role in students' course selection decisions, if such data were publicly available. The fact that evaluations are ranked third in importance behind course worth and grading leniency could be due to several factors. First, given that this information is not currently available at the university where this study was conducted, students may not have been able to gauge how important published evaluations would be in assisting with their course selection decisions. Second, the respondents were juniors and seniors enrolled in 300- and 400-level courses in their major, and previous research indicates that instructor attributes such as likability and enthusiasm typically captured in SET become less important as students progress to more advanced courses (Babad, Darley, and Kaplowitz 1999).

There is also the question of how students perceive the information content and credibility of teaching evaluations as defined in this study and published by most universities. Respondents were told that these evaluations represent base rate and consensus information derived from past student evaluations of teachers on ratings scales (e.g., 1 = *poor*, 3 = *excellent*). Comments by respondents[7] suggest that students may downgrade the credibility or usefulness of published course evaluations. As the student comments in parentheses after this sentence indicate, student do not trust the validity of the SET because they feel that the "other" students who complete them are very different from themselves or the typical student (e.g., "I don't trust course ratings because who knows what the students are like that did them?" and "Only 'A' students fill them out"; Borgida and Nisbett 1977); these atypical students do not provide useful informatino about the course and/or professor. Others failed to understand statistical inference and the predictive utility of large samples, giving more weight to a single individual's opinion of a teacher ("I'd rather talk to someone who has taken the class"; Tversky and Kahneman 1971). Still others mentioned that, consistent with previous research, SET only measure how "nice" or "popular" an instructor is and that they fail to give much information about other instructor or course attributes. Thus, one possible explanation for the lower importance weight assigned to SET is the perception that such information has low diagnostic value, that is, it does not provide very reliable

TABLE 5
THE EFFECT OF COURSE EVALUATIONS ON SHARE OF PREFERENCE FOR HYPOTHETICAL COURSE "PRODUCTS"

	Course "Products"		
Course Attributes	*Good Evaluations, Average on Other Attributes*	*Average on All Attributes*	*Poor Evaluations, Average on Other Attributes*
Course worth	Some useful knowledge	Some useful knowledge	Some useful knowledge
Grading leniency	Moderately easy grader	Moderately easy grader	Moderately easy grader
Course evaluations	Excellent	Average	Poor
Course workload	Moderate	Moderate	Moderate
Share of preference	58%	32%	10%

NOTE: Share of preference represents that percentage of the respondents who would prefer or choose each course "product," assuming these are the only three choices available. Shares of preference are ratio data.

TABLE 6
SHARE OF PREFERENCE FOR SIX HYPOTHETICAL COURSE "PRODUCTS"

	Course "Products"					
Course Attribute	*Average*	*High Course Worth but Low Evaluations*	*High Course Worth and Evaluations but Hard to Get Good Grades*	*Poor Evaluations but Easy Course*	*Good Evaluations but Low Course Worth*	*Good Evaluations and Little Work but Hard to Get Good Grades*
Course worth	Some useful knowledge	Great deal of useful knowledge	Great deal of useful knowledge	Some useful knowledge	Little useful knowledge	Some useful knowledge
Grading leniency	Moderately easy grader	Moderately easy grader	Very hard grader	Very easy grader	Moderately easy grader	Very hard grader
Course evaluations	Average	Poor	Excellent	Poor	Excellent	Excellent
Course workload	Moderate	Moderate	Heavy	Light	Moderate	Light
Share of preference	27%	26%	20%	15%	7%	5%

NOTE: Share of preference represents that percentage of the respondents who would prefer or choose each course "product," assuming these are the only six choices available. Shares of preference are ratio data.

information about a particular course (Dick, Chakrvarti, and Biehal 1990).[8]

The Importance of Perceived Course Worth in Course Choice

The fact that this attribute was found to be the most important factor in course choice for 3rd- and 4th-year business students is consistent with the business literature (e.g., Karns 1993; Kelley, Conant, and Smart 1991) and underlines the value these students place on gaining knowledge that is relevant to their major and career.[9] Surprisingly, students in this study consistently chose the course that would provide a "great deal" of useful knowledge even if the perceived workload was heavy and the instructor was believed to be a hard grader. While many may think of course worth as a function of the subject matter, it is clear from this study and previous research that students believe that the instructor plays a significant role in determining how relevant or useful the course is through his or her teaching effectiveness and pedagogical approach (Bacon and Novotny 2002; Karns 1993; Kelley, Conant, and Smart 1991).

This raises the interesting question of whether students are capable of assessing course worth. Students who may never have worked in the field for which they are preparing may not know how to distinguish "useful" from "useless" knowledge.[10] This possibility is given some credence by studies demonstrating that students' *self-perceptions* of academic performance are at variance with their *actual* learning and performance in a course, particularly for poorer performing students (Kennedy, Lawton, and Plumlee 2002; Moreland, Miller, and Laucka 1981). Judgments about course worth, therefore, may be biased and unreliable, and based on possi-

bly erroneous inferences. For example, business students may consider instructors who emphasize theory over application and real-world examples as providing relatively less useful knowledge. Similarly, students may positively associate course worth with instructors who possess extensive industry experience or who frequently use guest speakers in the classroom.

Clearly, students will not be able to identify a course that is in their best educational interests if they are unable to accurately assess course worth. Just as clearly, an instructor's pedagogical approach can influence students' perceptions of course worth (and therefore teaching evaluations; see Table 1). This is not to suggest that instructors restructure their courses to fit students' definitions of relevance or usefulness (see Armstrong [1995] for an interesting discussion of this point) but rather that faculty members become more proactive in educating and communicating to students why and how the knowledge gained in their courses is useful. Of course, there is the possibility that students' perceptions are correct in some instances and that business instructors may need to improve their course content to meet student needs and employer demands for usefulness and relevance. Peer evaluations of an instructor and course by academic colleagues, and in some cases by business practitioners, may provide valuable input to instructors concerned with improving course worth.

The Inverted-U Relationship between Perceived Workload and Course Choice

While student perceptions about workload had a statistically significant influence on course choice, this attribute was approximately five times less important in determining course choice than course worth and four times less important than grading leniency (Table 3). This relatively weak influence of perceived workload on course selection is consistent with previous research that has examined the effect of perceived workload on course choice (Babad, Darley, and Kaplowitz 1999; Coleman and McKeachie 1981). Of more interest is the finding that the "ideal" or most preferred course includes a moderate workload (Table 4). Preference for a moderate workload over a light one indicates that, up to a certain point, perceived workload and course worth may be positively related in students' minds. Several student comments lend support to this premise, for example, "If I hear that an instructor assigns a lot of work, I usually figure I will learn a lot of good stuff" and "How much work isn't as important as the kind of work the prof assigns; I like projects that make me apply the course material to the real world."

This inverted-U relationship between workload and course choice runs counter to previous SET studies that have reported either a positive (Bacon and Novotny 2002; d'Apollonia and Abrami 1997) or negative (Greenwald and Gillmore 1997) relationship between course workload and SET. However, the existence of a possible nonlinear relationship could not be explored in these studies, either because the authors chose to conduct correlational analyses, which assume linearity (e.g., Greenwald and Gillmore 1997) or because the workload factor included in the study had only two levels (low, high) (Bacon and Novotny 2002). The use of a three-level workload attribute in the present study affords more insight into students' workload preferences and suggests that the reported effects of workload on course selection or SET may depend on how a study is designed, that is, where the factor levels or workload scenarios fall on the inverted-U curve and how students interpret "heavy" and "light."[11]

Overall, these are encouraging findings for instructors who have felt pressured to lighten the workload in their courses in order to improve teaching evaluations and/or increase enrollment. If the workload is not too excessive, students perceive a positive relationship between workload and learning outcomes, as long as the assignments are believed to be useful and relevant to their major and career goals. On the other hand, the strong, positive relationship found between perceptions of an instructor's grading leniency and course choice suggests that a lenient grader who assigns a moderate workload will be preferred over a hard grader who assigns the same amount of work.[12]

STUDY LIMITATIONS

One limitation of the study is the problem with the independence of the SET attribute. While existing research does support the premise that SET primarily measure instructor popularity or student satisfaction, and respondents in this study mentioned this as well, future research needs to clarify for respondents exactly what the SET measure is meant to represent. Second, while the findings provide information about students' use of SET in course choice relative to other sources of information, the research design does not allow an assessment of the actual or absolute degree to which students would use SET in choosing courses. To ascertain the absolute importance of SET as an information source, the survey would need to include a measure of students' prior knowledge about the instructor. A third limitation relates to the generalizability of the findings: respondents attended one university, were primarily 3rd- and 4th-year students, and the reward for completing the survey (extra credit in most cases) created a self-selection bias. The hypothetical nature of the choice task also precludes any direct link between the findings reported here and the published SET field research using actual courses and teaching evaluations (e.g., Greenwald and Gillmore 1997).

While the choice-based conjoint analysis approach used in this study is an appropriate one given the study objectives, the part-worths or conjoint utilities derived from the data are at the aggregate level, precluding extensive segmentation analyses that directly test for segment differences in attribute and attribute-level importances. This is important given recent

findings demonstrating the moderating influence that individual-difference factors have on course choice and SET (e.g., Babad, Darley, and Kaplowitz 1999; Bacon and Novotny 2002). Recent software and hardware advances now make it possible to generate individual-level utilities from aggregate choice data using Hierarchical Bayes (Green and Krieger 2002), but this software was not available to the author during the time period when this research was conducted.

SUGGESTIONS FOR FUTURE RESEARCH

Need for Research on the SET–Course Choice Relationship

Clearly there is a need for more research on how SET influence course choice. This study focused on the role of official SET data in the *choice process*, given the trend toward publishing them online. Further research is needed to determine *actual usage* of online evaluations in course choice on campuses where these data are available; a search for usage statistics in preparation for this article yielded only anecdotal information. A replication of this study on campuses where official SET ratings are publicly available may change the relative importance of such evaluations, relative to the other attributes included here. Any such replication should also examine individual differences that may moderate the SET–course choice relationship, such as motivation, year in school, gender, and so on. As affordable Hierarchical Bayes (HB) software becomes more widespread, researchers will be able to combine CBC analysis and HB (CBC/HB) to model course preference functions for different student segments.

A second issue related to the SET–course choice relationship concerns the *format* of the SET information used in this study and reported by most universities: mean evaluation ratings (e.g., 1 = *poor*, 3 = *excellent*) from past students. Findings from previous research on the effects of concrete versus abstract information suggest that student guides containing student testimonials and other concrete information may exert a greater influence on course choice than the base rate and consensus information provided by average ratings (Babad, Darley, and Kaplowitz 1999; Borgida and Nisbett 1977; Dick, Chakrvarti, and Biehal 1990). Thus, an interesting question for future research is how the relative importance of SET in course choice may be influenced by the format in which they are presented. If faculty and administrators want to encourage students to use this information when selecting courses, it will be important to communicate the ratings in a format that is credible, understandable, and requires relatively low cognitive effort.

Need for Research on Course Choice and the Decision-Making Process

In light of the recent trend toward publishing SET online for use by students in selecting courses, more research is needed on the decision-making process preceding selection of a particular course. A greater understanding of the factors students consider and the heuristics they use to make course choices will allow administrators to develop tools to assist them in making optimal educational choices:

> Course selection (CS) decisions are too important to be made haphazardly, and college educators know that optimal CS contributes to better education, that relevant information must be provided to students in a style that will be acceptable and helpful to them. However, CS decision-making processes have not been studied intensively, and solid scientifically based information that will guide modes of practice are still lacking. (Babad, Darley, and Kaplowitz 1999, p. 167)

While the present study identifies and assesses the relative importance of several key instructor attributes, we know very little about how students *form* perceptions about an instructor's likability, workload, course worth, and grading leniency or how *accurate* these perceptions are. What sources of information do students use to assess course worth, for example? How important is word-of-mouth communication relative to other, more objective sources of information like course syllabi or SET? Is the decision process a compensatory one, or do students have a number of choice heuristics or decision rules they use? Several older studies have addressed the question of information source–using survey research, finding that "friends" are the chief source of information about instructors (e.g., Kerin, Harvey, and Crandall 1975; Martin 1989). However, a more qualitative approach that examines the motivations and rationales underlying course choice might offer some additional insights.

The question of how to assist business students in making course choices that optimize their educational experience has not been addressed empirically to the author's knowledge, other than the rare case study describing a self-help method or tool (e.g., "the course selection matrix chart"; Irving, Gorrell, and Johnson 1990). It may be instructive to review and apply the extensive literature on *career* choice to the study of course choice in the major, given the similar context or circumstances in which both decisions are typically made. For example, Gati (1986), in a review of the major issues in career decision making facing students when they graduate, identified four problems that could just as easily apply to the course choice context: (1) the unavailability of good information about alternative career options, (2) a lack of resources (time, effort) to collect all the information necessary for an informed decision, (3) the cognitive limitations of the decision maker, and (4) the lack of a framework for identifying and processing the relevant information. Gati (1986) and Gati and Asher (2001) have proposed a practical model of career choice based on Tversky's (1972) elimination-by-aspects theory of choice that offers a promising framework for the study and improvement of the course choice process. Such a model

(assuming empirical validation) could provide the conceptual foundation for the design of a Web-based application to assist students in making course choices,[13] similar to existing computer-assisted career guidance systems.

CONCLUSION

The present study identifies several key instructor attributes that are important in course choice for business majors and underscores the potentially important role that published, online SET may play in choice as they become available on more campuses. However, there is clearly a need for additional research aimed at improving our understanding of how students make course choices, with the goal of improving students' ability to select courses that will enhance their educational experience and further their educational objectives.

NOTES

1. See Haskell (1997a, 1997b, 1997d) for a review of a large body of legal rulings related to academic freedom and the use of student evaluations of teaching (SET) as evaluation tools by administrators. He views the current use of SET as an infringement on academic freedom.

2. A "none" option was not included in the present study because students do not typically have the option of not completing a particular required course in their major.

3. Given the design, the sample size (Respondents × Tasks) (800), and the number of concepts per task (2), ordinary least squares were used to make an approximation of the standard error associated with each main effect. This is then compared to the standard error for the "ideal" situation when the design is precisely orthogonal. The square of the ratio of the actual to the ideal standard error for each main effect gives the relative efficiency for this design.

4. Choice is a two-level nominally scaled variable, rather than an interval or ratio-scaled variable; hence the nonlinear or discontinuous nature of the relationship between marketing efforts and choice.

5. Sawtooths Software choice-based conjoint (CBC) software was used to conduct the logit analysis.

6. The randomized first-choice method (RFC) (Huber, Orme, and Miller 1999) was used to estimate shares of preference. It assumes the respondent will choose that product with the highest overall utility ("first-choice rule"), but it adds unique random error to the utilities. Each respondent is sampled many times to stabilize the share estimates. RFC also corrects for product similarity due to correlated sums of errors among products defined on many of the same attributes. The appropriateness of this method for the present study was validated with the holdout (fixed) task; RFC correctly estimated actual choice or preference within 3 percentage points.

7. Recall that there was an open-ended question at the end of the survey that asked respondents to describe their personal approach to selecting courses.

8. SET are also abstract, pallid data, relative to evaluations received directly from a fellow student. This may reduce the accessibility of the SET information in memory, which would then deflate the relative importance of the SET attribute at the time of choice (Dick, Chakrvarti, and Biehal 1990).

9. Responses to the open-ended question about personal approaches to course choice revealed that respondents interpreted "knowledge relevant to my major" as "knowledge that prepares me for the real world," rather than knowledge that would allow them to do better in a more advanced course in their major.

10. I am thankful to a reviewer of the article for raising this issue.

11. I am thankful to a reviewer of the article for suggesting this reason for the discrepancies between findings from this study and those reported in previous research. Bacon and Novotny (2002) did allude to the possibility of a nonlinear relationship between workload and SET in their article: "Workload preference is not a linear function but is instead curvilinear with some ideal point" (pp. 12-13).

12. This finding runs counter to the developmental study conducted by Babad, Darley, and Kaplowitz (1999) across many different college majors who found that perceptions of an instructor's grading leniency become less important in course selection as students progress to more advanced courses (these authors did not report results at the individual major level). It may be that business majors are concerned about their grades because they believe that potential employers use grade point average as a hiring criterion, a possibility that would be interesting to explore in future research.

13. The only published example of such a web application for course choice (to the author's knowledge) was published in an information technology journal and did not address the theoretical basis for including or excluding certain course and instructor attributes (Takeshita and Maeda 1999).

REFERENCES

Ahmadi, Mohammad, Marilyn Helms, and Farhad Raiszadeh. 2001. Business students' perceptions of faculty evaluations. *International Journal of Educational Management* 15 (1): 12-22.

Archibold, R. C. 1998. Payback time: Give me an "A" or else. *The New York Times Week in Review*, May 24, C1.

Armstrong, J. Scott. 1995. The devil's advocate responds to an MBA student's claim that research harms learning. *Journal of Marketing Education* 59 (3): 101-6.

Avi-Itzhak, Tamara. 1982. Teaching effectiveness as measured by student ratings and instructor self-evaluation. *Higher Education* 11:629-37.

Babad, Elisha, John M. Darley, and Henry Kaplowitz. 1999. Developmental aspects in students' course selection. *Journal of Educational Psychology* 91 (1): 157-68.

Bacon, Donald R., and Jenny Novotny. 2002. Exploring achievement striving as a moderator of the grading leniency effect. *Journal of Marketing Education* 24 (1): 4-14.

Barnoski, Robert P., and Alan L. Sockloff. 1976. A validation study of the faculty and course evaluation (FACE) instrument. *Educational and Psychological Measurement* 36:391-400.

Ben-Akiva, Moshe, and Steven R. Lerman. 1985. *Discrete choice analysis: Theory and application to travel demand.* Cambridge, MA: MIT Press.

Borgida, Eugene, and Richard Nisbett. 1977. The differential impact of abstract vs. concrete information on decisions. *Journal of Applied Social Psychology* 7:258-71.

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

Cahn, Steven. 1987. Faculty members should be evaluated by their peers, not by their students. *Chronicle of Higher Education*, October 14, B2.

Carroll, J. Douglas, and Paul E. Green. 1995. Psychometric methods in marketing research: Part 1, conjoint analysis. *Journal of Marketing Research* 32 (November): 385-91.

Chrzan, Keith, and Bryan Orme. 2000. An overview and comparison of design strategies for choice-based conjoint analysis. *Sawtooth Software Research Paper Series*. http://www.sawtoothsoftware.com/download/techpap/desgncbc.pdf (accessed May 20, 2003).

Clayson, Dennis E. 1999. Students' evaluation of teaching effectiveness: Some implications of stability. *Journal of Marketing Education* 21 (1): 68-75.

Coleman, Jeffrey, and Wilbert McKeachie. 1981. Effects of instructor/course evaluations on student course selection. *Journal of Educational Psychology* 73:224-26.

Costin, Frank, William T. Greenough, and Robert J. Menges. 1971. Student ratings of college teaching: reliability, validity, and usefulness. *Review of Educational Research* 41:511-36.

d'Apollonia, Sylvia, and Philip C. Abrami. 1997. Navigating student ratings of instruction. *American Psychologist* 52 (11): 1198-1208.

Davis, Richard, Misra Shekhar, and Stuart Van Auken. 2000. Relating pedagogical preference of marketing seniors and alumni to attitude toward the major. *Journal of Marketing Education* 22 (2): 147-54.

Deal, Ken. 2002. Get your conjoint online, in several flavors. *Marketing Research* 7 (Winter): 44-45.

Dick, Alan, Dipankar Chakrvarti, and Gabriel Biehal. 1990. Memory-based inferences during consumer choice. *Journal of Consumer Research* 17 (June): 82-93.

Feldman, Kenneth A. 1978. Course characteristics and college students' ratings of their teachers and courses: What we know and what we don't. *Research in Higher Education* 9:199-242.

———. 1989. The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multi-section validity studies. *Research in Higher Education* 30:583-645.

———. 1992. College students' views of male and female college teachers. Part 2: Evidence from students' evaluations of their classroom teachers. *Research in Higher Education* 33:415-74.

Freeman, Harvey R. 1994. Student evaluations of college instructors: Effects of type of course taught, instructor gender and gender role, and student gender. *Journal of Educational Psychology* 86:627-30.

Gati, Itamar. 1986. Making career decisions—A sequential elimination approach. *Journal of Counseling Psychology* 33 (4): 408-17.

Gati, Itamar, and Itay Asher. 2001. Prescreening, in-depth exploration, and choice: From decision theory to career counseling practice. *The Career Development Quarterly* 50 (December): 140-57.

Green, Paul E., and Abba M. Krieger. 2002. What's right with conjoint analysis? *Marketing Research* 7 (Spring): 25-27.

Green, Paul E., and V. Seenu Srinivasan. 1990. Conjoint analysis in marketing research: New developments and directions. *Journal of Marketing* 54 (October): 3-19.

Greenwald, Anthony G. 1997. Validity concerns and usefulness of student ratings of instruction. *American Psychologist* 52 (11): 1182-86.

Greenwald, Anthony G., and Gerald M. Gillmore. 1997. No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology* 89 (4): 743-51.

Haskell, Robert E. 1997a. Abridgement of academic freedom, promotion, reappointment and tenure rights by the administrative use of student evaluation of faculty: Part II, views from the court. *Education Policy Analysis Archives* 5 (17). http://epaa.asu.edu/epaa/v5n17.html (accessed September 10, 2002).

———. 1997b. Academic freedom, promotion, reappointment, tenure and the administrative use of student evaluation of faculty (SEF): Part III, analysis and implications of views from the court in relation to accuracy and psychometric validity. *Education Policy Analysis Archives* 5 (18). http://epaa.asu.edu/epaa/v5n18.html (accessed September 10, 2002).

———. 1997c. Academic freedom, promotion, reappointment, tenure and the administrative use of student evaluation of faculty (SEF): Part IV, analysis and implications of views from the court in relation to academic freedom, standards, and quality instruction. *Education Policy Analysis Archives* 5 (21). http://epaa.asu.edu/epaa/v5n21.html (accessed September 10, 2002).

———. 1997d. Academic freedom, tenure, and student evaluations of faculty: Galloping polls in the 21st century. *Education Policy Analysis Archives*, 5 (6). http://epaa.asu.edu/epaa/v5n6.html (accessed September 10, 2002).

Huber, Joel, Bryan Orme, and R. Miller. 1999. Dealing with product similarity in conjoint simulations. In *Sawtooth Software Conference Proceedings*. Sequim, WA: Sawtooth Software.

Irving, Eugene, Toni Gorrell, and Bruce Johnson. 1990. Helping students choose courses. *NACADA Journal* 10 (1): 51.

Johnson, Richard. 1996. Getting the most out of CBC: Part I and Part 2. *Sawtooth Software Technical Papers*. http://www.sawtoothsoftware.com (accessed August 14, 2002).

Karns, Gary. 1993. Marketing student perceptions of learning activities: Structure, preferences and effectiveness. *Journal of Marketing Education* 15 (1): 3-10.

Kelley, Craig A., Jeffrey S. Conant, and Denise T. Smart. 1991. Master teaching revisited: Pursuing excellence from the student's perspective. *Journal of Marketing Education* 13 (2): 1-10.

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): 243-52.

Kerin, Roger, Michael Harvey, and Nelson D. Crandall. 1975. Student course selection in non-requirement programs: An exploratory study. *Journal of Educational Research* 68:175-77.

Leventhal, Les, Phillip Abrami, and Raymond Perry. 1976. Do teacher rating forms reveal as much about students as about teachers? *Journal of Educational Psychology* 68:441-45.

Leventhal, Les, Phillip Abrami, Raymond Perry, and Lawrence Breen. 1975. Section selection in multi-section courses: Implications for the validation and use of teacher ratings forms. *Educational and Psychological Measurement* 35:885-95.

Louviere, Jordan D., David A. Hensher, and Joffre D. Swait, eds. 2000. *Stated choice methods: Analysis and application*. Cambridge, MA: Cambridge University Press.

Louviere, Jordan D., and George Woodworth. 1983. Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data. *Journal of Marketing Research* 20 (November): 350-67.

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 Michael Bailey. 1993. Multidimensionality of students' evaluations of teaching effectiveness: A profile analysis. *Journal of Higher Education* 64:1-18.

Marsh, Herbert W., and Dennis Hocevar. 1991. The multidimensionality of students' evaluations of teaching effectiveness: The generality of factor structures across academic discipline, instructor level, and course level. *Teaching and Teacher Education* 7:9-18.

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

Martin, Elaine S. 1989. The relationship among college student characteristics and their assigned importance to dimensions of course choice information. *Journal of College Student Development* 30:69-76.

McKeachie, Wilbert J. 1997. Student ratings: The validity of use. *American Psychologist* 52 (11): 1218-25.

Moreland, Richard, John Miller, and Frances Laucka. 1981. Academic achievement and self-evaluation and metacomprehension ability. *Journal of Educational Psychology* 73 (3): 335-44.

Nichols, Alan, and John C. Soper. 1972. Economic man in the classroom. *Journal of Political Economy* 80:1069-73.

Orme, Bryan. 2002. Conjoint analysis has value. *Marketing Research* 14 (Winter): 46-47.

Ory, John C., and Katherine Ryan. 2001. How do student ratings measure up to a new validity framework? *New Directions for Institutional Research* 109:27-44.

Pohlmann, John T. 1975. A description of teaching effectiveness as measured by student ratings. *Journal of Educational Measurement* 12:49-54.

Radmacher, Sally A., and David J. Martin. 2001. Identifying significant predictors of student evaluations of faculty through hierarchical regression analysis. *Journal of Psychology* 135 (3): 259-68.

Takeshita, T., and K. Maeda. 1999. An integrated Web computing application for tasks related to course selection and registration. *Information and Software Technology* 41:995-1004.

Tarleton, Alice. 2003. A blunt instrument: In an appeal to consumer power the government will ask students to rate the teaching on their courses as a guide to applicants—to the alarm of lecturers. *The Guardian* (Manchester, UK), March 4, D2.

Theall, Michal, and Jennifer Franklin. 2001. Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction? *New Directions for Institutional Research* 109:45-56.

Tversky, Amos. 1972. Elimination by aspects: A theory of choice. *Psychological Review* 79:281-99.

Tversky, Amos, and Daniel Kahneman. 1971. Belief in the law of small numbers. *Psychological Bulletin* 6:105-10.

Vence, Deborah L. 2003. Companies look to tools that improve sites, connect goals. *Marketing News*, May 12, 4.

Vriens, Marco, Harmen Oppewal, and Michel Wedel. 1998. Ratings-based versus choice-based latent class conjoint models—An empirical comparison. *Journal of the Market Research Society* 40 (1): 237-48.

Williams, Wendy M., and Stephen J. Ceci. 1997. How'm I doing? Problems with student ratings of instructors and courses. *Change*, September–October, 13-23.

Wilson, Robin. 1998. New research casts doubt on value of student evaluations of professors. *The Chronicle of Higher Education*, January 16, A12-A14.