

2004

## Course Selection Decisions by Students on Campuses With and Without Published Teaching Evaluations

Wendy Bryce Wilhelm

Charles Comegys

Follow this and additional works at: <https://scholarworks.umass.edu/pare>

---

### Recommended Citation

Wilhelm, Wendy Bryce and Comegys, Charles (2004) "Course Selection Decisions by Students on Campuses With and Without Published Teaching Evaluations," *Practical Assessment, Research, and Evaluation*: Vol. 9 , Article 16.

DOI: <https://doi.org/10.7275/axmj-zh11>

Available at: <https://scholarworks.umass.edu/pare/vol9/iss1/16>

This Article is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Practical Assessment, Research, and Evaluation by an authorized editor of ScholarWorks@UMass Amherst. For more information, please contact [scholarworks@library.umass.edu](mailto:scholarworks@library.umass.edu).

# Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited. PARE has the right to authorize third party reproduction of this article in print, electronic and database forms.

Volume 9 Number 16, August 2004

ISSN 1531-7714

## Course Selection Decisions by Students on Campuses With and Without Published Teaching Evaluations

Wendy Bryce Wilhelm, *Western Washington University* &  
Charles Comegys, *Merrimack College*

In spite of students' increasingly vocal demands for access to official student evaluations of teaching (SET), little is known about the relative importance of SET in course selection decisions, and whether such evaluations are viewed by students as a valuable source of information about an instructor or course. Using conjoint analysis and a web survey to assess SET importance, we found that business students on campuses with published SET rated course evaluations as *less important* in course choice than students on campuses without published evaluations. Moreover, student perceptions of the amount of useful knowledge gained in the course and how lenient the instructor is in his/her grading practices were found to have the greatest influence on course choice within the business major.

College students in the U.S. first began evaluating faculty in 1926, but it was not until the 1960s that student evaluations of instructors' teaching effectiveness began to be formally initiated on many campuses (d'Apollonia and Abrami, 1997). Today, 90-100% of colleges and universities across the U.S. engage in this practice (Trout, 2000).

The primary purpose of student evaluations of teaching (SET) is to provide faculty with feedback to assist them in improving instructional quality. SET are also heavily used by administrators when making personnel decisions involving tenure and promotion (Haskell, 1997a; Marsh, 1987). A third critical user group is students, who may use SET, when publicly available, to help them select which courses and instructors to take.

In spite of students' increasingly vocal demands for access to official SET (Foster, 2003; Tarleton, 2003), little is known about the relative importance of SET in course selection decisions, and whether such evaluations are viewed by students as a valuable source of information about an instructor or course. The course

selection process is an important area of investigation because of the serious impact course choices have on the overall quality of and students' satisfaction with the education received, and on the career direction students take.

The present study replicates and extends a recent conjoint study conducted by one of the authors that examined the relative influence or importance of SET and other instructor attributes on business students' preference for a set of hypothetical courses in their major. The present study extends this original study -- which only surveyed students from one university that does not publish their SET in any form -- to business students from several U.S. universities that vary with respect to the availability of published, online SET. This larger and more diverse sample allowed us to examine the relative influence of SET on course choice for students who do have access to published SET versus those who do not enjoy such access.

We first review the existing research on course choice and state the research question investigated in this empirical study. We then describe the methodology used -- choice-based conjoint analysis -- and the study

findings. The paper concludes with a discussion of the implications the findings have for understanding and improving the course choice process and several limitations that reduce the generalizability of the findings.

## LITERATURE REVIEW and RESEARCH QUESTION

### Validity of SET as a Measure of Teaching Effectiveness

Because of their widespread use and influence on promotion and tenure decisions, it is not surprising that in higher education the most prevalent area of research has revolved around the question of whether SET are valid measures of teaching effectiveness. Well over 2000 articles having been written on the topic (Wilson, 1998). Some researchers report that student evaluations are generally statistically reliable and valid predictors of overall teaching effectiveness (Braskamp, Brandenburg and Ory, 1984; Marsh, 1984; Whitworth, Price and Randall, 2002), while some suggest that SET are primarily a measure of instructor popularity (Marks, 2000) or a measure of how hard/lenient the instructor's grading practices are (Greenwald and Gillmore, 1997). The current controversy over the validity of SET may have a negative impact on students' perceptions and use of evaluations in course choice.

### The Course Selection Process

#### *Complexity of the Course Selection Process.*

Selection of the 'right' course(s) may be described as a high involvement, high risk decision-making situation because the cumulative effect of the series of choices students make each semester/quarter may impact their college major selection, their ability to take additional course work, as well as their career direction and future employment opportunities. There are a plethora of factors that students may consider in their course selection decisions as they choose between competing and attractive course alternatives, including perceptions about a course's workload, the instructor's grading leniency, the usefulness of the knowledge gained in the course, the instructor's reputation, and the times/days the course meets. According to Babad, Darley and Kaplowitz (1999): "In course selection, not one, but multiple, sequential and interdependent decisions must

be made concurrently. The projected utilities are sometimes contradictory. . . and different courses are selected with different objectives in mind" (p. 157).

When a student's objective is to select a course in his/her major that is taught by more than one instructor, it is reasonable to expect that more time and effort will be expended in order to assure a satisfactory outcome. This is confirmed by Babad et al. (1999): "Students reported their decisions about different courses are based on different considerations, with most serious thought being devoted to selecting among courses within their major field of study for upperclassmen, and to deciding on courses that might help them test out a possible major for underclassmen. It is on decisions about those (primary) courses that the students expend most thought, and come closest to the optimum of rational decision making" (p. 167). A rational decision making process might also include a search for a heuristic or highly credible information source to simplify course selection decisions.

*The Role of SET in Course Choice.* There are many sources of information available to assist students in selecting a course. These include college bulletins, academic advisors, course descriptions, course syllabi, student published course guides or Web sites, informal word of mouth, and official, published SET. With respect to making official SET available to students, many colleges and universities are currently debating whether to publish evaluations of teaching effectiveness (Babad, Darley & Kaplowitz, 1999).

Coleman and McKeachie (1981) found that instructor/course evaluations had an impact on student selection of courses. Their results showed that students choose the highest rated course in spite of its reportedly heavy workload. Several studies have found that faculty reputation influences student course selection. In a study involving section selection in multi-section courses, faculty reputation was found to be a primary reason for section choice, and the most frequently cited source of instructor reputation information was reports from other students (Leventhal, Abrami, Perry and Breen, 1975). Borgida and Nisbett (1977) found that brief, face-to-face comments from students influenced course selection. Further, they concluded that statistical student rating data had little impact on the course selection decision. This finding is consistent with those of several other studies that have reported that students

prefer more concrete, anecdotal course information over student evaluation data collected by formal, university-sanctioned instruments (Borgida, 1978; Coleman and McKeachie, 1981; Hendel, 1982).

In general, there appears to be some ambivalence surrounding the usefulness of SET in course choice, with student-produced guides and word-of-mouth frequently preferred over SET as an information source about an instructor's teaching ability. However, most of these studies were conducted in the 1970s and 1980s, and their findings did not distinguish between students who had had direct experience with using SET in course choice versus those who had not. It is conceivable that students who have had access to and have used published SET to make course choices over time may feel more or less positive about the diagnostic value of SET than students who have had no direct experience with them.

### **Student Demand for and Availability of Published SET**

Students across the country are now highly interested in making SET available online (Haskell, 1997b; Tarleton, 2003). A recent survey of students to determine their level of interest in published student ratings of instruction concluded that students favor published ratings of instruction and rate the likelihood of potential benefits from published evaluations as high (Howell and Symbaluk, 2001). Numerous colleges and universities have responded to this call by publishing their formal faculty evaluation data on-line. Other institutions have not prevailed in court when they attempted to deny student access to SET (Haskell, 1997b, note 55).

Widespread student demand for instructor and course evaluation feedback online for use in making informed course/instructor selections is further evidenced by the recent emergence of Internet sites such as: [pick-a-prof.com](http://pick-a-prof.com), [professorperformance.com](http://professorperformance.com), [ratingsonline.com](http://ratingsonline.com), [RateMyProfessor.com](http://RateMyProfessor.com), [ratingsonline.com](http://ratingsonline.com), [reviewum.com](http://reviewum.com), [whototake.com](http://whototake.com) and [teacherreviews.com](http://teacherreviews.com). A recent article in the *Chronicle of Higher Education* states: "Students at . . . colleges are increasingly seeking electronic access to their classmates' evaluations of professors. When administrators at some institutions fail to meet this demand, Pick-A-Prof often

swoops in to woo student-government leaders" (Foster, 2003, p. A33).

These sites generally present a compilation of informal and anonymous student reviews and comments on faculty and courses, describe professional quirks, and present testing and grading patterns (Lewin, 2003). Lawsuits have been filed against several of these sites claiming defamation and intentional infliction of distress (Anonymous, 2000; Carlson, 2000; Fisher, 2001). Resolution of the debate on how the First Amendment applies to Internet speech with respect to potentially libelous and slanderous postings on such sites will most likely have to be decided in the courts.

### **Validity of Students' Internet Site Evaluations of Professors as a Measure of Teaching Effectiveness**

Research on the validity of the information students record online concerning teaching effectiveness is lacking. However, the limited evidence suggests that online SET are primarily a measure of a professor's popularity, findings similar to those reported by Greenwald and Gillmore (1997) and others. For example, a recent study by Felton, Mitchell and Stinson (2003) suggests that students' high-quality ratings of their professors posted on [RateMyProfessors.com](http://RateMyProfessors.com) may not be a valid measure of teaching effectiveness because these data are significantly influenced by other factors. These authors concluded that the instructor's appearance and how easy he or she makes a course play a role in students' ratings of their professor's quality of teaching.

The questionable validity and reliability of the instructor ratings provided by such online sites suggests that university administrators might do well to develop their own, potentially more valid SET instruments and make them publicly available to all enrolled students. Such an investment on the part of universities requires evidence that SET are a useful and important tool in course choice.

### **Replication and Extension of Original Study**

In 2003, the first author investigated the relative influence of published SET, grading leniency, course workload, and course worth (whether the faculty member provides useful knowledge relevant to the

student's major) on hypothetical course choice within a student's major (Wilhelm 2004). The selection of these key attributes was based on a review of the literature and on several pretests with students. The study, involving undergraduate third and fourth year business majors from an institution that does not publish SET, revealed that course worth, grading leniency and published SET were the most important factors influencing course choice or preference. These findings are consistent with earlier studies that concluded that SET information plays a key role in course selection but is not necessarily the most significant factor considered in the student's decision making process (Borgida, 1978; Borgida and Nisbett, 1977; Coleman and McKeachie, 1981; Hendel, 1982; Leventhal *et. al.*, 1975). However, as noted earlier, the generalizability of these findings is limited by the fact that most respondents had no real-world experience using actual SET to make course choices.

## Research Question

The present study *replicates* the original study in an effort to further our understanding of which attributes most influence student preference for a particular course, and *extends* that research in an important way by surveying students from universities that publish SET online as well as students from universities that do not publish SET. If student experience with published SET has been positive (negative), then we might expect SET to exert more (less) influence on course choice, relative to the influence reported by students on campuses without published SET.

**Research Question: Will the SET attribute be perceived as a more or less important influence on course choice by students from campuses where SET are published online versus students from campuses that do not publish SET online?**

## METHODOLOGY

### Research Design

Choice-based conjoint analysis was used to assess students' preferences for various hypothetical courses that varied with respect to several instructor attributes. Sawtooth Software's CBC System was used to conduct a full profile conjoint analysis study. An in-depth description of the research methodology and

analyses used in this study can be found in Appendix 1, excerpted from Wilhelm (2004).

Four instructor attributes were included in the choice study, based on previous research that indicates that they are key attributes students use when choosing a course/section from among other required courses in their major to enroll in for a particular quarter. These attributes are: course evaluations (official SET), grading leniency, course workload and whether the instructor provides useful knowledge relevant to the student's major. Each of the attributes had three levels (see Table 1).

**TABLE 1: Attributes and Attribute Levels Used In Conjoint Task**

Course Evaluations	Grading Leniency	Course Workload	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

A fractional factorial, randomized experimental design was used to generate an optimal set of concepts to present to each respondent. [1] The randomization is done by the conjoint software as part of the experimental design process, so that each respondent receives a unique series of conjoint questions or tasks; thus there were 193 different surveys, equal to the number of respondents. 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, i.e., the two products presented were the same for all respondents. One of the fixed choice tasks was placed first and served as a "practice" question (i.e., the data were not used in statistical analyses). The second fixed task was inserted in the middle of the randomized choice tasks, serving as a holdout task to provide an indication of how well the utility data generated from the randomized tasks would 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. The actual instructions to the respondents and an example of a choice task are presented in Figure 1. [2] Within each choice task, the presentation order of the attributes was randomized; in other words, the course evaluation attribute was not always presented first, as it is in Figure 1.

contact with faculty at schools across the U.S. Of the 54 faculty who were contacted, 39% or 21 agreed to have their students participate; rates of acceptance were not significantly different for the two sets of schools (published versus unpublished SET). [3] All faculty were given the same instruction sheet to read to their students. The instructions asked students for their assistance, described the purpose of the study and provided the URL where students could access the survey (see Figure 2 for the complete instructions).

**FIGURE 1: Example of Choice Task**

If these were the only course section options available for a particular required course in your major, which one 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):

Professor and Course receive <b>average student ratings</b> , as published on the WEB	Professor and Course receive <b>excellent student ratings</b> , as published on the WEB
<b>Very difficult</b> to get an "A" or "B" in this Professor's course	<b>Very easy</b> to get an "A" or "B" in this Professor's course
<b>Light workload</b> assigned by Professor	<b>Heavy workload</b> assigned by Professor
Professor provides <b>little useful knowledge</b> relevant to my major	Professor provides a <b>great deal of useful knowledge</b> relevant to my major
<input type="radio"/>	<input type="radio"/>

Some instructors gave their students extra credit to participate, some made participation mandatory, and some just asked students to complete the conjoint survey. All surveys were completed online, outside of the classroom, at students' convenience.

**FIGURE 2: Student Instruction Sheet**

**WE NEED YOUR HELP!**

Would you please complete an important survey which should take you no more than 15-20 minutes? Your responses will remain completely anonymous.

We are interested in finding out what factors you consider when you are deciding which particular section of a required course in your major/concentration to enroll in for a particular quarter or semester. For example, if there were four sections of a required course offered next quarter by different instructors, what causes you to prefer one section over another? We realize that scheduling (the days/times each section is offered) has a significant influence on your selection decision, but for this study we want you to assume that ALL sections are offered at days and times that are convenient for you.

Information about what factors influence your decision will help faculty and administrators to design course and section schedules that better reflect students' desires. Findings from this study will also reveal whether we need to provide more and/or different information to students about each section of a course (e.g., each instructor's past course evaluations in this course), in order to help students decide which section to enroll in.

Your participation in this study is essential and is greatly appreciated. Please take this study seriously, and answer each question honestly and completely.

**To begin the survey, go to:** <http://www.cbe.wvu.edu/survey/cour/cou2logn.htm>

Thank you again for your help!

Following the choice tasks, respondents answered several questions about their (potential) use of published course evaluations in selecting courses, what they thought course evaluations measured, and what sources of information they typically used to decide on specific courses to take in their major. All data were collected online, and the survey instrument can be accessed at: <http://www.cbe.wvu.edu/survey/cou2/cou2logn.htm>

**Sample and Procedure**

A sample of the population of interest – students at four year universities that did or did not publish teaching evaluations on-line -- was obtained through personal

## RESULTS

### Sample Characteristics

A total of 193 respondents completed the web survey, 129 from schools that do not publish their official course evaluations online and 64 from schools that do publish them. Sixty-nine of the completed surveys came from students at a small private school that does not publish their course evaluations online. These respondents were not included in the analyses because (1) their inclusion would create problems with subgroup analyses due to unequal cell sizes, and (2) their inclusion could create a possible confound when interpreting the findings, due to the fact that they came from a different type of university than the rest of the sample. The findings reported below are based on a usable sample of 124 -- 60 from schools that do not publish their SET, and 64 from schools that do.

The “published SET” group included respondents from two large, public universities in the Western states, while the “unpublished SET” group came from five large and mid-sized public universities across the U.S. There were no statistically significant differences between the two respondent groups in gender, year in school or average GPA. Respondents were primarily female (63%), undergraduate juniors or seniors majoring in Marketing (95%), with an average GPA of 3.24 (s.d. = .5). Sample characteristics for each respondent group and the overall sample are summarized in Table 2.

### Sources of Information about Courses (non-conjoint questions)

There were significant differences between the published SET group and the unpublished SET group (hereafter PUB and UNPUB) on several of the survey questions dealing with their use of SET and other sources of information in course choice. Table 2 shows that almost half of the UNPUB group said they would “never use” SET to assist them in course selection decisions, if they were available. This is not surprising, given these students lack of familiarity with such a tool. However, one third of the PUB group also said they would never use SET, and only 16 percent said they would always use them, implying that these students may not have found this tool very helpful in course choice. Further, when asked to select the important sources of information they used to assist them in course

choice, only 3 percent of the PUB group said that they reviewed published SET.

The PUB group’s apparent lack of confidence in SET as a diagnostic tool (in absolute terms and relative to the UNPUB group) is confirmed by this group’s belief that SET do not really communicate much about how effective a teacher is, only how much students liked him/her (58% of the PUB group felt that SET provided these two types of information, versus 74% of the UNPUB group). Both groups agreed that they generally gave professors they liked higher ratings on course evaluations (mean = 4.15, s.d. = .97, 1= strongly disagree, 5= strongly agree).

### Analysis of Conjoint Data: Logit Model

The choice data were analyzed using multinomial logit analysis (MNL). Logit was chosen because the form of the dependent and independent variables is categorical. 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. [4] These weights are analogous to “importance weights” 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). Hierarchical Bayes estimation techniques were then applied to the aggregate level part-worths generated from the logit analysis in order to obtain individual level utilities.[5] Individual level utilities are necessary if the analysis plan calls for subgroup comparisons, as ours did (i.e., a comparison of respondents who attended universities that published their official course evaluations on-line with those who attended universities that did not make evaluations available to students).

Both main and interaction effects models were examined to determine which model best fit the data. The best model included all main effects plus one 2-way interaction term: Course Evaluations X Grading Leniency (chi-square tests indicated that all main effects and this interaction term were statistically significant, i.e., significantly affected course choice). The addition of

**TABLE 2: Sample Characteristics and Course Choice Information Sources**

Characteristic/Question	Overall (n=124)	Published SET (n=64)	Unpublished SET (n=60)
Gender (M/F) <sup>1</sup>	37% / 63%	32% / 68%	43% / 57%
Year in School <sup>1</sup>	60% seniors 35% juniors	60% seniors 34% juniors	60% seniors 36% juniors
Average GPA <sup>1</sup>	3.24 (s.d. = .50)	3.40 (s.d. = .30)	3.08 (s.d. = .52)
Use of published SET if available <sup>2</sup>	49.5% (would) never use 10.5% (would) always use	31.3% never use 15.6% always use	48.3% would never use 5.0% would always use
Type of Information provided by SET <sup>2</sup>	1. Whether students liked this professor (48%) 2. How much work there will be in the course (19%) 3. whether the instructor is an effective teacher (16%) 4. whether this course will be useful for my major/career (11%)	1. Whether students liked this professor (45%) 2. How much work there will be in the course (25%) 3. whether the instructor is an effective teacher (12.5%) 4. whether this course will be useful for my major/career (12.5%)	1. Whether students liked this professor (52%) 2. whether the instructor is an effective teacher (20%) 3. how much work there will be in the course (13%) 4. whether this course will be useful for my major/career (10%)
Sources of Information used to assist in course choice <sup>2</sup>	1. student testimonials and/or Student Guide (71.4%) 2. faculty advisor/business professors (22%) 3. check course info - syllabus, web site, description (3%) 4. review published, on-line SET (2%)	1. student testimonials and/or Student Guide (63%) 2. faculty advisor/business professors (28%) 3. check course info - syllabus, web site, description (6%) 4. review published, on-line SET (3%)	1. student testimonials and/or Student Guide (82%) 2. faculty advisor/business professors (15%) 3. check course info - syllabus, web site, description (3%) 4. review published, on-line SET (0%)
“I generally give professors who I like higher ratings on course evaluations.” (1=strongly disagree, 5=strongly agree)	4.15 (s.d. = .97)	4.24 (s.d. = .88)	4.14 (s.d. = .95)

<sup>1</sup> No statistically significant differences were found between respondents from schools with published versus unpublished SET.

<sup>2</sup> Statistically significant differences (chi square tests; p < .05) exist between respondents from published SET schools and those from schools that do not published SETs.

this interaction term significantly increased the explanatory power of the model as determined by a chi-square test between the main effects-only model and a second model including the interaction term. [6]

**Relative Attribute and Attribute Level Importance**

The relative importance of each instructor attribute in course choice is presented in Table 3. While each attribute had a statistically significant influence on choice, no significant differences in importance weights were found between the two respondent

groups. Perceptions about the knowledge to be gained in a course and the leniency of the instructor’s grading policy were the two most important attributes. Directionally, the PUB group places less importance on course evaluations and more importance on course worth (knowledge gained) than the UNPUB group.

**TABLE 3: Relative Attribute Importance**<sup>1</sup>

Attribute	Overall (n=124)	Published SET (n=64)	Unpublished SET (n=60)
	Relative Importance (Chi-square, p value)	Relative Importance (Chi-square, p value)	Relative Importance (Chi-square, p value)
Knowledge Gained In Course	33% (129.48, p < .01) <sup>2</sup>	37% (75.39, p < .01)	30 % (58.86, p < .01)
Grading Leniency	29% (106.62, p < .01)	26% (40.85, p < .01)	32% (67.96, p < .01)
Course Evaluations <sup>3</sup>	24% (65.69, p < .01)	21% (24.91, p < .01)	26% (42.46, p < .01)
Course Workload	14% (25.43, p < .01)	16% (15.63, p < .01)	12% (10.06, p < .01)

<sup>1</sup> The relative importance of each attribute is calculated by computing the difference between the largest and smallest part-worth for each attribute, summing the differences, and normalizing to 100.

<sup>2</sup> The chi-square test determines whether an attribute plays a significant role in respondents' choice of course section (degrees of freedom = 2 for all tests). Note that, for each of the three groups, all four attributes are statistically significant.

<sup>3</sup> The difference in the importance of course evaluations in choice between respondents from published SET schools (21%) and those from schools that do not publish SET (26%) is not statistically significant. None of the between group chi-square tests revealed a statistically significant difference in attribute weights between these two groups of respondents.

What does the ideal instructor/course look like? Table 4 ranks the attribute levels from most preferred to least preferred for the total sample; there were no statistically significant differences in the rank orders between the two respondent groups. Not

surprisingly, the ideal instructor would provide a great deal of useful knowledge, assign a light workload, be a lenient grader and receive excellent course evaluations. Findings for both the ideal product configuration and the attribute importance rankings are

**TABLE 4: Ranking of Attribute Levels Based on Average Utility Values**<sup>1</sup> (n= 124)<sup>2</sup>

Rank	Course Attributes (Utilities)			
	Course Worth (avg. utility value)	Grading Leniency (avg. utility value)	Course Evaluations (avg. utility value)	Course Workload (avg. utility value)
1	A Great Deal of Useful Knowledge (64.21)	Very Easy to get an "A/B" (49.73)	Excellent (44.47)	Light (24.73)
2	Some Useful Knowledge (-3.05)	Moderately Easy to get an "A/B" (21.34)	Average (0.42)	Moderate (7.61)
3	Little Useful Knowledge (-61.16)	Very Difficult to get an "A/B" (-71.07)	Poor (-44.89)	Heavy (-32.34)

<sup>1</sup> 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., light workload and moderate grading leniency) are equally preferred.

<sup>2</sup> There are no statistically significant differences in the ranking of attribute levels between the two groups of respondents (those from published SET schools, those from schools where SET are not published); thus, only the attribute level rankings for the overall sample are presented here. These represent an average of the two respondent groups' rankings.

**TABLE 5: The Effect of Course Evaluations on Share of Preference<sup>1</sup> for Hypothetical Course “Products”**

Course “Products”			
Course Attributes	Excellent 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
<b>Share of Preference</b>			
<b>Published SET:</b>	<b>84%</b>	<b>15%</b>	<b>1%</b>
<b>Unpublished SET:</b>	<b>91%</b>	<b>9%</b>	<b>0%</b>

<sup>1</sup> Share of Preference represents that percent 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.

similar to the findings reported in the original study, with the exception that a moderate workload was preferred over a light one in the first study.

**Share of Preference for Different Course Configurations**

The part-worths derived from the logit and HB analyses were used to simulate market conditions that present a hypothetical mix of course “products” from which to choose in any given quarter/semester. Sawtooth Software’s Market Simulator was used to run the simulations, which produce “share of preference” or market share data for each hypothetical course, assuming these were the only courses from which to choose.<sup>[7]</sup> Tables 5 and 6 describe two possible market scenarios of interest.

All else being equal, Table 5 shows that the UNPUB group of students is ten times more likely (91% versus 9%) to choose a course/section with an instructor that receives excellent, as opposed to average, course evaluations; the PUB group is only five to six times more likely to do so (84% versus 15%). Similar to the findings discussed above from the non-conjoint survey questions, respondents at schools where course evaluations are published on-line place *less importance* on excellent evaluations than do those respondents who do not have access to evaluations. Given the demands being made by many students to require official course/faculty evaluation results to be published on-line so that they can use these results to select “good” courses, it is not surprising that the findings show that the UNPUB group believes that SET results would be a helpful tool in course choice. However, if they truly are

helpful, one would expect that the PUB group would have a *greater* preference for a course with excellent course evaluations than the UNPUB group, not a lower one as the findings suggest.

Table 6 shows a more complex market, with five potential course “products” to choose from. What trade-offs are respondents willing to make? Here we can see quite clearly that, for both groups, course evaluations are not the most important factor determining course selection (see Table 3). The share of preference data show that the UNPUB group chooses courses that will most likely get them a good grade and that don’t have too much work. The PUB group, on the other hand, appears to be most concerned about course worth (knowledge gained) regardless of the grading leniency of the instructor or the workload assigned.

**DISCUSSION AND IMPLICATIONS**

**Student Use of SET in Course Choice**

The findings that the PUB group placed less importance on course evaluations (relative to the UNPUB group) and was not particularly influenced by improvements in course evaluations (e.g. from average to excellent) when selecting a preferred course suggest that students may believe SET will be useful in course choice --- until they actually get a chance to use them for that purpose. The low SET usage rates reported by students on campuses where the ratings are publicly available suggest that these students have not found the

**TABLE 6: Share of Preference<sup>1</sup> for Six Hypothetical Course “Products”**

Course Attributes	Course “Products”				
	High Course Worth but Low Evaluations	High Course Worth & Evaluations but Hard to get a Good Grade	Poor Evaluations but Easy Course	Good Evaluations but Low Course Worth	Good Evaluations and Little Work But Hard to Get a Good Grade
Course Worth	Great Deal of Useful Knowledge	Great Deal of Useful Knowledge	Some Useful Knowledge	Little Useful Knowledge	Some Useful Knowledge
Grading Leniency	Moderately Easy Grader	Very Hard Grader	Very Easy Grader	Moderately Easy Grader	Very Hard Grader
Course Evaluations	Poor	Excellent	Poor	Excellent	Excellent
Course Workload	Moderate	Heavy	Light	Moderate	Light
<b>Share of Preference</b>					
<b>Published SET:</b>	<b>25%</b>	<b>31%</b>	<b>21%</b>	<b>12%</b>	<b>10%</b>
<b>Unpublished SET:</b>	<b>26%</b>	<b>21%</b>	<b>26%</b>	<b>14%</b>	<b>13%</b>

<sup>1</sup> Share of Preference represents that percent 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.

ratings to be particularly helpful in course choice. Why is this?

One possible explanation has to do with the validity of the SET instrument itself. As mentioned earlier, considerable controversy exists over what SET actually measure. If they are primarily a measure of popularity and are easily manipulated by doing “popular” things in class (e.g., showing lots of videos, using entertaining guest speakers), then students (particularly the better ones who want to learn something) may not put much faith in their predictive validity. The fact that both the PUB and the UNPUB groups in this study strongly agreed that they give professors whom they like higher ratings on course evaluations demonstrates the positive relationship between “instructor liking” and teaching evaluations. Validity and reliability can also be affected by how students complete the ratings scales: “If students have no faith in the system and put little thought and effort into their evaluations, then, regardless

of the sophistication of the techniques used to test the validity of evaluation results, the results will be useless” (Marlin, 1987, p. 715). Clearly more work needs to be done to validate SET instruments or at least improve students’ perceptions of their validity.

A second possible explanation for the PUB group’s apparent disillusionment with SET as a tool for improving the course choice process may have to do with the attributes or questions typically included in the instrument. For example, one often sees a question pertaining to the instructor’s record for coming to class on time. There is nothing in the literature to suggest that this is an important consideration in course choice (although it may have diagnostic value when it comes to evaluating faculty for T&P). Perhaps a separate instrument needs to be devised by and for students that includes measures of such attributes as “knowledge provided by the instructor that is relevant to my major,” (where 1 = none, 5 = a great deal). Take grading

leniency as another example. Many students desire grade distribution information; in our study grading leniency had a significantly more important influence on course choice than SET -- for both the PUB and the UNPUB groups. According to Foster (2003): "...some university Web services that feature student evaluations, like those at Austin and Penn State, disappoint students by not posting the grade distributions of professors" (p. A33). Are grade distributions a valid measure of grading leniency? Should grade distribution data be included as part of an institution's online SET published information? These questions deserve further study and discussion.

Third, perhaps the typical format of published SET -- statistical ratings and consensus base rate information -- causes students some difficulty due to its level of abstraction and numerical form. Borgida and Nisbett (1977) found that statistical student rating information had little impact on course selection, while brief and vivid face-to-face comments from others had a much greater impact. It is clearly important to present the SET data in a format that students are likely to benefit from. Faculty and administrators may want to explore other formats for reporting SET information (e.g., graphs).

### Study Limitations and the Need for Replication

It is essential to acknowledge the limitations of this study that reduce the generalizability of the findings. First, the sample was limited to only a few schools. Second, the PUB sample was composed of students from two large public universities in the western U.S., while the UNPUB sample was drawn primarily from middle-sized, public universities. While there is no evidence to suggest that students at large universities would respond differently from those at smaller schools, future research should replicate this study with a larger sample of universities and colleges.

Third, the importance of each of the course attributes may be differentially affected by subgroup differences on variables such as level of intrinsic interest in the course material, importance of maintaining/improving average GPA, or whether a student is employed or not. [8] These differences should be examined in future research by including questions about students' motives for choosing a course in their major and other individual difference variables

such as number of hours/week employed and whether they receive financial aid or not (aid is often contingent upon GPA).

## CONCLUSION

There is a critical need for further research on the course choice process. A greater understanding of course choice may assist faculty and administrators in the development of decision support systems that will help students to make better choices and thus lead to greater student satisfaction with the educational experience. It is hoped that the concepts and findings discussed in this initial empirical study will, if nothing else, increase researchers' awareness of the many aspects of the course choice process that remain to be explored.

## REFERENCES

- Anonymous (2000). Class Action: 'Trashed by Students' Online Teacher Ratings, A San Francisco Professor Fights Back. *People Weekly*, 53 (24), 163-164.
- Babad, Elisha, John M. Darley and Henry Kaplowitz (1999). Developmental Aspects in Students' Course Selection. *Journal of Educational Psychology*, 91 (1), 157-168.
- Ben-Akiva, Moshe, and Steven R. Lerman (1985). *Discrete choice analysis: theory and application to travel demand*. Cambridge: The MIT Press.
- Borgida, E. (1978). Scientific Education -- Evidence is not necessarily Informative: A Reply to Wells and Harvey. *Journal of Personality and Social Psychology*, 36, 477-482.
- Borgida, E. and R. Nisbett (1977). The Differential Impact of Abstract vs. Concrete Information on Decisions. *Journal of Applied Social Psychology*, 7, 258-271.
- Braskamp, L.A., D.C. Brandenburg and J.C. Ory (1984). *Evaluating Teaching Effectiveness: A Practical Guide*. Beverly Hills, California: Sage.
- Carlson, Scott (2000). Two Instructors Sue a College over a Web Site's Anonymous Teacher Evaluations. *Chronicle of Higher Education*, 46 (41), A41.
- Coleman, Jeffrey and W. J. McKeachie (1981). Effects of Instructor/Course Evaluations on Student Course Selection. *Journal of Educational Psychology*, 73 (2), 224-226.
- D'Appollonia, S. and P.C. Abrami (1997). Navigating Student Ratings of Instruction. *American Psychologist*, 52 (11), 1198-1208

- Felton, James, John Mitchell and Michael Stinson (2003). Web-Based Student Evaluations of Professors: The Relations between Perceived Quality, Easiness, and Sexiness. Working Paper Series, Dept. of Finance and Law, Central Michigan University.
- Fisher, Marla Jo (2001). Instructor-Review Web Site Again Comes Under Lawsuit Threats. *Community College Week*, 13 (13), 9.
- Foster, Andrea L. (2003). Picking Apart Pick-A-Prof: Does the Popular Online Service Help Students Find Good Professors, or Just Easy A's? *Chronicle of Higher Education*, 49 (26), A33.
- Greenwald, Anthony 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), 743-751.
- Haskell, R.E. (1997a). Academic Freedom, Tenure, and Student Evaluations of Faculty: Galloping Polls in the 21<sup>st</sup> Century. *Education Policy Analysis Archives*, 5(6). <http://epaa.asu.edu/epaa/v5n21.html> (accessed January 10, 2004).
- Haskell, R.E. (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 January 10, 2004).
- Hendel, P. (1982). Evaluating the Effects of a Course Evaluation System Designed to Assess Students in Electing Courses. Paper presented at the *Annual Meeting of the American Educational Research Association*.
- Howell, Andrew J. and Diane G. Symbaluk (2001). Published Student Ratings of Instruction: Revealing and Reconciling the Views of Students and Faculty. *Journal of Educational Psychology*, 93 (4), 790-797.
- Huber, Joel, Bryan Orme, and R. Miller (1999). Dealing with product similarity in conjoint simulations. *Sawtooth Software Conference Proceedings*, Sequim WA: Sawtooth Software.
- Johnson, Richard (1996). Getting the Most out of CBC: Part I and Part 2. *Sawtooth Software Technical Papers*. Retrieved 14 August 2002 from <http://www.sawtoothsoftware.com>.
- Leventhal, Les, Philip C. Abrami, Raymond P. Perry, and Lawrence J. Breen (1975). Section Selection in Multi-Section Courses: Implications for the Validation and Use of Teacher Rating Forms. *Educational and Psychological Measurement*, 35, 885-895.
- Lewin, Tamar (2003). New Online Guides Allow College Students to Grade Their Professors. *New York Times*, March 24<sup>th</sup>, Section A, 11.
- Marks, Ronald B. 2000. Determinants of student evaluations of global measures of Instructor and course value. *Journal of Marketing Education* 22 (2), 108-119.
- Marlin, James W., Jr. (1987). Student Perception of End-of-Course Evaluations. *Journal of Higher Education*, 58 (6), 704 – 716.
- Marsh, H.W. (1984). Students' Evaluation of University Teaching: Dimensionality, Reliability, Validity, Potential Biases, and Utility. *Journal of Educational Psychology*, 27, 707-754
- Marsh, H.W. (1987). Students' Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research. *International Journal of Educational Research*, 11, 253-387.
- Seldin, P. (1984). *Changing Practices in Faculty Evaluation*. San Francisco: Josey-Bass.
- Tarleton, A. (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, B6.
- Trout, Paul (2000). Flunking the Test: The Dismal Record of Student Evaluations. *Academe*, 86(4), 58-61.
- Whitworth, James E., Barbara A. Price and Cindy H. Randall (2002). Factors that Affect College of Business Student Opinion of Teaching and Learning. *Journal of Education for Business*, 77 (5), 282-290.
- Wilhelm, Wendy (2004). The Relative Influence of Published Teaching Evaluations and Other Instructor Attributes on Course Choice. *Journal of Marketing Education* 26 (1), 17-30.
- Wilson, Robin (1998). New Research Casts Doubt on Value of Student Evaluations of Professors. *The Chronicle of Higher Education*, 44 (19), A12-A15.

---

## Appendix I

### Description of the Conjoint Model and Methodology Used in the Present Study

Excerpted from Wilhelm (2004), pp. 20-22.

Students' stated preferences for course options were evaluated using conjoint analysis. Conjoint 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 pre-specified 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/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 by marketing research practitioners has experienced significant growth (relative to ratings-based conjoint analysis) as "more companies want to understand how people make *choices*?" (Vence, 2003, p. 4, emphasis added). 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, i.e., 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, 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, i.e., 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 choice-based conjoint analysis, based on a multinomial logit model of choice (Louviere et al., 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 et al., 1998). For these reasons, we utilized 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 in both studies.

**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 sixty business majors was conducted to confirm the importance of the attributes identified by previous research as being potentially the most important in course choice 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 (1-5 scale), (2) the importance of any additional attributes they identified (1-5 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 X 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 1 are the most important ones in choice of a required course section.

Each of the attributes used in the two studies had three levels (low, moderate, high; see Table 1). 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 1 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 et al., 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.

## Appendix I References

Carroll, J. Douglas, and Paul E. Green (1995). Psychometric methods in marketing research: Part 1, conjoint

- analysis. *Journal of Marketing Research* 32 (November), 385-391.
- Chrzan, Keith, and Bryan Orme (2000). An overview and comparison of design strategies for choice-based conjoint analysis. *Sawtooth Software Research Paper Series*. Retrieved 20 May 2003 from <http://www.sawtoothsoftware.com/download/techpap/desgncbc.pdf>.
- Deal, Ken (2002). Get your conjoint online, in several flavors. *Marketing Research* 7 (Winter), 44-45.
- 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.
- Johnson, Richard (1996). Getting the Most out of CBC: Part I and Part 2. *Sawtooth Software Technical Papers*. Retrieved 14 August 2002 from <http://www.sawtoothsoftware.com>.
- Louviere, Jordan D., David A. Hensher, and Joffre D. Swait, eds (2000). *Stated Choice Methods: Analysis and Application*. Cambridge: 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.
- Orme, Bryan (2002). Conjoint analysis has value. *Marketing Research* Winter, 46-47.
- Vence, Deborah L (2003). Companies look to tools that improve sites, connect goals. *Marketing News* 12 May, 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.
- Wilhelm, Wendy (2004). The Relative Influence of Published Teaching Evaluations and Other Instructor Attributes on Course Choice. *Journal of Marketing Education* 26 (1), 17-30.

---

## Appendix II

### Background Information on Conjoint Analysis (Sawtooth Software)

- Choice-Based Conjoint (CBC) Technical Paper, [http://sawtoothsoftware.com/download/techpap/cbc\\_tech.pdf](http://sawtoothsoftware.com/download/techpap/cbc_tech.pdf)
- Understanding Conjoint Analysis in 15 Minutes, <http://sawtoothsoftware.com/download/techpap/undca15.pdf>
- An Overview and Comparison of Design Strategies for Choice-Based Conjoint, <http://sawtoothsoftware.com/download/techpap/desgncbc.pdf>
- For a list of all of Sawtooth's conjoint-related technical papers, see the *Technical Paper Series*, Sawtooth Software, <http://sawtoothsoftware.com/techpap.shtml>
- 

### Notes:

[1] This is explained in more depth in Appendix 1, under **Experimental design and dependent measure**. Briefly, the particular experimental design approach used here to generate the fractional factorial (but nearly orthogonal) design is the balanced overlap method. This method employs random sampling with replacement for choosing concepts, permitting some level overlap within the same task. For more information on CBC design strategies, and justification for using a fractional factorial design, see <http://sawtoothsoftware.com/download/techpap/desgncbc.pdf>.

[2] A “none” option was not included in the study since students do not typically have the option of not completing a particular required course in their major.

[3] As noted earlier, we were interested in comparing schools who publish their official course evaluation ratings data online for each instructor/course/quarter versus those schools that do not publish them in any form. Online evaluation results at the former schools (we could only identify 25 such schools) are available to anyone with a University ID (see, for example, the University of Washington's "course evaluation catalogue", described at <http://www.washington.edu/oea/uwusers.htm>, the last paragraph on "public access.")

[4] Sawtooth Software choice-based conjoint (CBC) software was used to conduct the logit analysis.

[5] Sawtooth Software HB software was used to generate the individual level utilities. HB can significantly improve upon aggregate models such as logit for conjoint/choice analysis or any other situation in which respondents provide multiple observations. By using HB estimation, researchers can improve the reliability and predictive validity of their models. Two technical papers, available at <http://sawtoothsoftware.com/download/techpap/undhb.pdf> and <http://sawtoothsoftware.com/download/techpap/hbwhy.pdf>, provide a basic overview of HB estimation and explain why this statistical technique is currently receiving so much attention from researchers.

[6] There was a significant improvement in RLH and log-likelihood when the one interaction term was added (relative to main effects model; the change in LL vs. Main Effects Model = 9.5, chi-square = 19,  $p < .01$ ; RLH=93%).

[7] 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.

[8] Thanks to an anonymous reviewer for suggesting this possible confound and future research idea.

### Citation:

Wilhelm, W.B, & C. Comegys (2004). Course Selection Decisions by Students on Campuses With and Without Published Teaching Evaluations. *Practical Assessment, Research & Evaluation*, 9(16). Available online: <http://pareonline.net/getvn.asp?v=9&n=16>

### Authors

Wendy Bryce Wilhelm  
Associate Professor of Marketing  
College of Business and Economics  
Western Washington University  
Bellingham WA 98225-9073  
(360) 650-4816  
(360) 650-4844 (fax)  
Wendy.Wilhelm [at] wwu.edu

Charles Comegys  
Associate Professor of Marketing  
Girard School of Business & International  
Commerce  
Merrimack College  
North Andover MA  
(978) 837-5409  
(978) 837-5086 (fax)  
ccomegys [at] merrimack.edu