Journal of Agricultural and Applied Economics, 37,3(December 2005):635–645

Measuring the Impact of Externalities on College of Agriculture Teaching Evaluations

Ronald A. Fleming, Ernest F. Bazen, and Michael E. Wetzstein

Student evaluation of teaching (SET) is employed as an aid in improving instruction and determining faculty teaching effectiveness. However, economic theory indicates the existence of externalities in SET scores that directly influence their interpretation. As a test of this existence, a multinomial-choice, ordered data estimation procedure is employed to identify course externalities influencing SET. These externalities include student class standing, required courses, class size, days a class meets, class meeting time, classroom location, and classroom design. Results indicate that externalities have a significant impact on teaching evaluations. Thus, failure to internalize these externalities will lead to biases in SET and questionable use of SET as an aid in instruction improvement and determining faculty effectiveness.

Key Words: externalities, ordered probit, SET, teaching evaluation

JEL Classifications: A20, A22, I21

At most Land Grant universities, student evaluation of teaching (SET) plays a role in determining faculty effectiveness (Baker et al.; Seevers et al.; Worley and Casavant). While SET data can be used to improve instruction or a course (Frey; Worley and Casavant), there are critics who feel that SET transforms classroom instruction into a popularity contest and that the instructor's role is changed from that of educator to entertainer (Baker et al.; Becker and Watts; Greenwald; Greenwald and Gillmore; Wilson). For these reasons, arguments that the SET has led to a decline in education standards cannot be ignored (Becker and Watts).

Students, regularly present in the classroom, appear uniquely situated to provide first-hand information concerning teaching (Frey). Yet a student's perception may not encompass the full set of criteria for quality instruction. At the University of Kentucky, faculty members contend that administrators base a large percentage of faculty teaching effectiveness on SET. If SET for an instructor rises, this is viewed as a necessary and sufficient condition for improved instruction. However, the link between SET and instruction effec-

---

R.A. Fleming is associate professor, Department of Agricultural Economics, University of Kentucky, Lexington, KY. E. Bazen is assistant professor, Department of Agricultural Economics, University of Tennessee, Knoxville, TN. M.E. Wetzstein is professor, Department of Agricultural and Applied Economics, University of Georgia, Athens, GA.

The authors wish to thank David Debertin, professor, and Harry Hall, professor emeritus, Department of Agricultural Economics, University of Kentucky, and two anonymous reviewers of the *Journal* for their useful comments and insights. This is Technical Paper 05-04-072 of the Kentucky Agricultural Experiment Station.

tiveness may not hold when the full set of criteria for quality instruction has been taken into account. SET might well be a very weak signal for teaching effectiveness.

The literature is rich in studies concerning how best to interpret the results of SET and their validity (see Becker and Watts; Seevers et al.). Yet limited studies have focused on externalities impacting SET scores. As the term suggests, teaching faculty have limited direct influence on these externalities. Whether the course is required, day of the week, and time the course is offered, location of classroom, and type of classroom (i.e., a fully automated "smart" classroom versus a classroom equipped with a chalkboard and overhead projector) are some of the externalities involved. The influence of these externalities on SET has indirectly received only a limited investigation.

Braskamp et al. (1983, 1984) evaluated the impact on SET of course level, class size, and undergraduate versus graduate students. Following Braskamp et al., Baker, Rudd, and Hoover measured the correlation between students' final grade, class size, course level, time of evaluation, required versus elective course, and SET. A weakness of these previous studies is their limited investigation of SET in terms of an instructor's overall welfare and the use of simple correlations and linear regressions for estimating the statistical relationships between endogenous factors under an instructor's control and course externalities.

While sharing an interest in some of the externalities that might affect SET scores, this study is a fundamental departure from previous literature. Specifically, the objectives of this research are to employ economic theory for identifying relevant components of SET and, based on this theory, construct testable hypotheses. A further objective is to demonstrate the appropriate empirical methodology for investigating statistical relationships involving SET. The focus of this investigation is SET scores of faculty within the College of Agriculture (CA) at the University of Kentucky (UK). This study utilizes individual student records from each course offered by the college from the fall semester of 1997 to the fall semester of 2002, excluding summers (11 semesters representing 26,519 potential observations).

Theoretical Model

Consider the following separable utility function, U, so a faculty member follows the utility maximizing paradigm of maximizing his or her scholarly program

$$(1) \quad U(\vec{X}, N, \vec{E}, \vec{R}),$$

where $\vec{X}$ is a vector of scholarly activities a faculty member has control over, N is the faculty's SET score, $\vec{E}$ is a vector containing a faculty's effort levels toward instruction, and $\vec{R}$ denotes a vector of a faculty member's course rigor. It is hypothesized faculty receive a positive response from scholarly activities, $\vec{X}$, SET scores, N, and providing rigor in their courses, $\vec{R}$. Exerting effort toward instruction, $\vec{E}$, is hypothesized to negatively impact satisfaction.

In the attempt to maximize Equation (1), an instructor is constrained by limited resources (one notable limit is time) represented by the following implicit production possibilities frontier:

$$(2) \quad F[\vec{X}, N(\vec{E}, \vec{R}, \vec{Z})],$$

where $\vec{Z}$ is the vector of externalities associated with SET scores. These externalities could augment or diminish the scores depending on whether the externality is positive or negative. Through resource allocation, an instructor can enhance scholarly activities and SET scores, but on the efficiency frontier, there is a tradeoff in scholarly activities for SET scores. A faculty's effort toward instruction, $\vec{E}$, is hypothesized to positively influence SET scores, N, whereas, course rigor, $\vec{R}$, will dampen them.

The first-order conditions from maximizing Equation (1) subject to Equation (2), given λ as the Lagrangian multiplier, are

$$\frac{\partial U}{\partial \vec{X}} - \lambda \frac{\partial F}{\partial \vec{X}} = 0,$$

$$\frac{\partial U}{\partial \vec{E}} + \frac{\partial U}{\partial N}\frac{\partial N}{\partial \vec{E}} - \lambda\frac{\partial F}{\partial N}\frac{\partial N}{\partial \vec{E}} = 0,$$

$$\frac{\partial U}{\partial \vec{R}} + \frac{\partial U}{\partial N}\frac{\partial N}{\partial \vec{R}} - \lambda\frac{\partial F}{\partial N}\frac{\partial N}{\partial \vec{R}} = 0,$$

$$\frac{\partial U}{\partial N}\frac{\partial N}{\partial \vec{Z}} - \lambda\frac{\partial F}{\partial N}\frac{\partial N}{\partial \vec{Z}} = 0.$$

Solving for the Lagrangian multiplier in the above first-order conditions yields

$$(3) \quad \lambda = \frac{\partial U}{\partial \vec{X}}\Big/\frac{\partial F}{\partial \vec{X}} = \left(\frac{\partial U}{\partial \vec{E}} + \frac{\partial U}{\partial N}\frac{\partial N}{\partial \vec{E}}\right)\Big/\left(\frac{\partial F}{\partial N}\frac{\partial N}{\partial \vec{E}}\right) = \left(\frac{\partial U}{\partial \vec{R}} + \frac{\partial U}{\partial N}\frac{\partial N}{\partial \vec{R}}\right)\Big/\left(\frac{\partial F}{\partial N}\frac{\partial N}{\partial \vec{R}}\right) = \frac{\partial U}{\partial N}\Big/\frac{\partial F}{\partial N}.$$

At the optimum, the ratio of marginal benefits to marginal cost is the same across all the determinants. In terms of effort, $\vec{E}$, the marginal benefits of SET score enhancement from increased effort, $(\partial U/\partial N)/(\partial N/\partial \vec{E}) > 0$, is tempered by the direct loss in marginal benefits $\partial U/\partial \vec{E} < 0$. For rigor, $\vec{R}$, the effect is just the opposite. The loss in marginal benefits from increased rigor through the SET score, N, $(\partial U/\partial N)/(\partial N/\partial \vec{R}) < 0$, tempers the direct gain, $\partial U/\partial \vec{R} > 0$. As modeled, the externalities, $\vec{Z}$, do not influence the first-order conditions. The last term in Equation (3) is independent of $\vec{Z}$. A faculty member will not modify his or her levels of scholarly activities, effort, or rigor in the face of differing course externalities. However, the total SET score received, which is influenced by course externalities, does directly impact satisfaction, yielding the condition $(\partial U/\partial N)/(\partial F)/(\partial N)$. The key to these first-order conditions is the influence effort, $\vec{E}$, rigor, $\vec{R}$, and externalities, $\vec{Z}$, have on SET scores, N. Thus, the empirical question flowing from this theoretical development is whether the hypothesized signs on these variables are supported empirically. A corollary question is: If they are supported, what is the magnitude of these effects?

Empirical Investigation

In this investigation, individual student SET records obtained from the UK Office of Institutional Research (www.uky.edu/IR/) were employed to define N, $\vec{E}$, $\vec{R}$, and $\vec{Z}$. Given the ordinal nature of SET scores, N, ordered probit estimation was used to test the hypotheses and to determine the magnitude of the associated coefficients.

Students at UK are asked to respond to SET questions by penciling in the appropriate bubble on a form that is optically scanned and electronically recorded. For most questions, students are asked to indicate if they strongly disagree, disagree, agree, or strongly agree with the stated question. In the case of questions "Overall value of the course," and "Overall quality of teaching," the options are poor, fair, good, or excellent. Table 1 reports the 19 questions asked in Sections B, C, D, and E of UK's SET form.

Individual student records from fall semester 1997 to fall 2002, where course evaluations were conducted in undergraduate UK College of Agriculture classes, are included in the data set (11 semesters; 26,519 individual records/observations). In addition to instructor and course name, the SET form elicits information on class standing, expected grade, reason for taking the course, hours spent weekly studying for the class, size of the class, the type of instructor (e.g., regular, full-time faculty, or graduate teaching assistant), and the type of class (i.e., lecture versus lab). Gender and race demographics are not provided. Secondary class information was collected to augment the data provided by the SET form. Program brochures were reviewed to determine what classes in each program are required of all students in the program. Many students mark a class as required if advised to take the class even if it is not a required class according to university guidelines. Information was also collected on what days of the week the class met, the grade-point average for the class measured at the end of the semester, the time of day that the class met, and the location of the room in which the class met.

Summary statistics for this data set arranged by elements of the vectors $\vec{E}$, $\vec{R}$, and $\vec{Z}$ are provided in Table 2. As indicated in the table, the dependent variable, N, is composed of the ordinal response to the question "Overall quality of teaching." The design of the

Table 1. Questions on University of Kentucky's SET form[a]

Question Number	Question
Section B—course items	
1	Outlined course material and grading
2	Textbook contributed to understanding
3	Supplemental readings and assignments helped understanding
4	Exams reflected what was taught
5	Grading was fair and consistent
6	Assignments were distributed evenly
7	Graded assignments returned promptly
8	Graded assignments included comments
Section C—instructor items	
9	Presented material effectively
10	Had good knowledge of the subject matter
11	Was available for consultation
12	Satisfactorily answered class questions
13	Stimulated interest of the subject
14	Encouraged class participation
Section D—learning outcomes	
15	Learned to respect different viewpoints
16	Increased my ability to analyze and evaluate
17	Course helped my ability to solve problems
18	Gained understanding of concepts and principals
19	Course stimulated me to read further
Section E—summary items	
20	Overall value of the course
21	Overall quality of teaching

[a] Students are asked to indicate if they strongly disagree, disagree, agree, or strongly agree with questions 1 through 19. For questions 20 and 21, students are asked to indicate poor, fair, good, or excellent.

evaluation form is such that students draw on their previous responses to make what is, in effect, a final judgment. The score for N is used by faculty and administrators alike as a final assessment of teaching ability.

For the effort vector, $\vec{E}$, the questions on UK's SET form that indicate a faculty member's effort in teaching activities include "exams reflected what was taught," "graded assignments returned promptly," "graded assignments included comments, presented material effectively," and "was available for consultation" (Table 2). Recall that the underlying hypothesis is that faculty who invest more time in instructional effort will positively influence SET scores, N.

The elements of $\vec{R}$ are, perhaps, the most difficult to measure. As indicated in Table 2, courses where students agree that the course improved their ability to analyze, evaluate, and solve problems are deemed more rigorous courses. More rigorous courses are also associated with increased study time, a lower class grade-point average (GPA; measured after the course was completed), and a lower expected letter grade in the class. Students indicate the time spent per week studying for the course on the SET form by selecting the appropriate time range. In the ordered probit estimation, each time range is treated as a separate dichotomous variable. Students are also asked to indicate what letter grade they expect to receive in the class. The class GPA and expected-grade variables could be difficult to interpret. The hypothesis to be tested is if rigorous classes, as measured by a lower class GPA and lower expected grade, result in lower SET scores, N.

Table 2. Summary Statistics: Mean, Standard Deviation, Data Range, and Difference Between the Mean of the Included and Excluded Observations[a]

Variable	Expected Sign	Mean	Standard Deviation	Data Range	Difference in Mean
Dependent variable					
Overall quality of teaching, N		3.474	0.744	1–4	0.084
Teaching effort, $\tilde{E}$					
Exams reflect what was taught	+	3.382	0.645	1–4	0.014
Graded assignments returned promptly	+	3.350	0.681	1–4	0.109
Graded assignments included comments	+	3.271	0.686	1–4	0.047
Presented material effectively	+	3.363	0.713	1–4	0.060
Was available for consultation	+	3.486	0.576	1–4	−0.017
Class rigor $\tilde{R}$					
Increased my ability to analyze and evaluate	+	3.315	0.643	1–4	−0.020
Course helped ability to solve problems	+	3.253	0.663	1–4	−0.003
Studied 1 or fewer hours per week	+	0.275	0.446	0–1	0.057
Studied 2 hours per week	+	0.318	0.466	0–1	0.120
Studied 3 hours per week	+	0.238	0.426	0–1	0.085
Studied 4–5 hours per week	−	0.119	0.323	0–1	0.007
Studied 6–7 hours per week	−	0.032	0.175	0–1	−0.018
Class grade-point average	−	3.041	0.463	1.2–4.0	−0.257
Grade student expected to earn	+	3.335	0.724	0–4	−0.209
Class externalities, $\tilde{Z}$					
College standing, freshman	−	0.122	0.327	0–1	0.068
College standing, sophomore	−	0.131	0.338	0–1	0.038
College standing, junior	+	0.284	0.451	0–1	0.133
College standing, senior	+	0.454	0.498	0–1	0.156
Class required for major	±	0.634	0.482	0–1	0.128
Class reported as required	−	0.215	0.411	0–1	−0.022
Freshman in a required class	±	0.097	0.296	0–1	0.054
Sophomore in a required class	±	0.099	0.299	0–1	0.029
Junior in a required class	±	0.183	0.387	0–1	0.085
Senior in a required class	±	0.249	0.433	0–1	0.074
Freshman in a class reported as required	±	0.013	0.113	0–1	0.007
Sophomore in a class reported as required	±	0.018	0.133	0–1	0.006
Junior in a class reported as required	±	0.070	0.255	0–1	0.039
Senior in a class reported as required	±	0.113	0.317	0–1	0.055
Class enrollment	−	36.420	21.978	4–132	10.055
Class meets Monday, Wednesday, and Friday	−	0.402	0.490	0–1	0.092
Class meets Tuesdays and Thursdays	+	0.409	0.492	0–1	0.074
Class meets twice during the week	−	0.077	0.266	0–1	−0.014
Class meets once during the week	−	0.111	0.314	0–1	−0.085
Class meets between 8 and 10 a.m.	−	0.274	0.446	0–1	0.037
Class meets between 10 a.m. and 12 p.m.	+	0.326	0.469	0–1	0.096
Class meets between 12 and 2 p.m.	+	0.238	0.426	0–1	0.017
Class meets between 2 and 4 p.m.	−	0.130	0.336	0–1	−0.020
Class not located on the College of Agriculture campus	−	0.166	0.372	0–1	0.039
Modern, smart classroom	+	0.672	0.470	0–1	0.178
Classroom not liked by faculty	−	0.219	0.414	0–1	0.110

[a] Included are 14,394 of 26,519 University of Kentucky College of Agriculture SET observations.

The externalities, $\vec{Z}$ include the maturity of the student; if a class is required; the size of the class; when the class meets during the week; time of day that the class meets; and classroom characteristics, including location and design (Table 2). Each element of $\vec{Z}$ is hypothesized to contribute to a faculty's SET score, *N*, yet each is beyond his or her direct control. Note that other endogenous variables were included in the analysis but are not discussed. These independent variables control for variation in SET scores, *N*, that is common to department, program, semester, and year (i.e., the fixed effects in the data). In terms of hypothesis testing, Table 2 lists the elements of the externalities vector $\vec{Z}$ along with their associated expected signs.

Maturity is measured by class standing (i.e., freshman, sophomore, junior, or senior). In this case, maturity better reflects a student's understanding of the university and career objectives than it does of one's age. Separate dichotomous variables are used to represent each student classification. It is hypothesized freshmen and sophomores, being new to the college life, see little connection between course content and real-world realities. On the other hand, juniors and seniors are taking classes in their chosen major and tend to appreciate the connection between course content and interests. The associated negative signs for freshmen and sophomores and positive signs for juniors and seniors reflect these hypotheses.

True required courses are those that are taken by all students seeking a particular major. At UK, these courses are designated core courses. But students are also advised strongly to take certain classes that are needed to fulfill a total-hour requirement within a major and/or to fulfill emphasis area requirements. As indicated in Table 2, it is hypothesized that a course required of all students in order to obtain a degree is associated with lower SET scores, while the sign on major required courses is indeterminant. At UK, core courses that stress understanding of concepts and principles tend to be sophomore- and junior-level courses. As such, they are more tedious. Major required courses tend to be junior- and senior-level classes within a student's emphasis area. The tedious concepts and principles may be offset by student interest in the major.

There is a possible interactive relationship between maturity and required courses. At UK, core courses are often taken in the sophomore year. Thus, it is difficult to distinguish if it is a lack of maturity or the fact that the course is required that is tied to a lower SET score. To test if there is a relationship between maturity and required courses, each of the student classification variables is multiplied by the variables representing the two types of required classes (eight variables in total). It is not possible to determine *a priori* how the different interaction terms will affect SET rating because increased maturity is anticipated to have a positive impact, while taking a required course is anticipated to have a negative impact.

Class size and associated SET scores are a common theme in the literature. The literature generally indicates faculty prefer smaller classes given the relative ease of teaching and managing. Thus, class size is included as a continuous variable, although evaluation records from the same class will share a common value for a number of students. It is anticipated that the SET score and class size are inversely related.

Four dichotomous variables were developed related to day of the week that the class meets. The first retains the value of one if the class meets Monday, Wednesday, and Friday and is zero otherwise. The second retains the value of one if the class meets Tuesday-Thursday and is zero otherwise. The third variable retains the value of one if the class meets twice during the week and is zero otherwise. Monday-Wednesday and Wednesday-Friday are the most common twice-a-week offerings, but other combinations are possible. The fourth variable retains the value of one if the class meets once during the week (any day of the week) and is zero otherwise. The literature is silent on the subject of how meeting day impacts SET scores. However, making a connection between longer classes and student lack of focus, then longer classes would lower SET scores. Following this logic, classes that meet once a week are anticipated to be associated with lower SET scores relative to classes that meet three times a week.

Unlike day of the week, there are clear expectations as to how time of day impacts SET scores. Instructors that hold classes during the optimal 10 a.m. to 2 p.m. window are thought to be more favorably assessed. Four dummy variables are derived to assess the impact that time of day has on SET scores. If the class start time is within the scope of the variable, that variable retains the value of one and is zero otherwise. The variables representing the 8–10 a.m. and 2–4 p.m. time slots are expected to have a more negative impact on SET score than do the variables representing the 10 a.m. to 12 p.m. and 12–2 p.m. time slots.

The final three externality variables capture the effect on SET scores associated with student satisfaction concerning physical characteristics of the classroom. Physical characteristics include location, if the classroom is designated as smart, and if the classroom is generally disliked by faculty who teach there.

The majority of CA classes are taught in buildings generally described as the CA campus. For classes not on the CA campus, travel cost to class becomes a factor. To test the impact of location on SET score, a dichotomous variable is used to identify classes held in the CA. It is anticipated that classes located away from the CA campus will act to reduce SET scores.

Over the past several years, the CA has invested significant resources toward upgrading classrooms. Upgrades included computer consoles, LCD projectors, ELMOs, VCRs, DVDs, and electronic blackboard capabilities. These upgrades are designed to improve instruction and learning. A dichotomous variable is used to distinguish between CA campus classrooms that have been upgraded (the smart classrooms) and those that have not. Given that the classrooms were upgraded to improve instruction, it is hypothesized that being assigned to a smart classroom will have a positive impact on SET scores.

Finally, with respect to physical classroom characteristics, there are some classrooms in which faculty dislike teaching due to relatively poor layout. A dichotomous variable is used to identify classes held in these unsuitable classrooms. If the characteristics of a classroom are such that it is difficult to teach, then it will also be difficult to learn. Although the classroom is the problem, it is anticipated that students will respond to learning difficulties by lowering their SET scores for quality of teaching.

Results

The ordinal nature of SET scores, N, indicates the appropriate estimation procedure is ordered logit or ordered probit analysis (Greene). Linear regression (OLS) would treat the difference between responses of poor and fair the same as the difference between responses of good and excellent. Multinomial logit or probit, on the other hand, would fail to account for the ordinal nature of N.

The underlying functional form associated with ordered probit estimation is nonlinear, so the parameter estimates cannot be interpreted as the change in the dependent variable resulting from a one-unit change in an independent variable. Proper interpretation requires estimation of the marginal effects for individual estimates.

The dependent variable SET scores, N, 14 variables associated with teaching effort, $\vec{E}$, and class rigor, $\vec{R}$, along with the 26 externalities, $\vec{Z}$, are employed in the ordered probit estimation. Of the available 26,519 individual student records, 14,394 records contained complete information. Many students declined to provide information on expected course grade and/or number of hours studied per week, thus eliminating those records from the sample. Tests of samples comprised of the included and excluded observations reveal that the mean and variances of the two samples are statistically different ($\alpha = 0.05$). However, while the differences are statistically significant, they are not empirically significant (Table 2).

The results of the ordered probit estimation are reported in Table 3. Comparison of the log-likelihood values between the estimated model (−7709.97) and a model restricted to

Table 3. Ordered Probit Estimation Results

Variable	Parameter Estimate[a]	Standard Error
Intercept	−6.031*	0.435
Threshold parameter 1	1.603*	0.044
Threshold parameter 2	3.570*	0.050
Teaching effort, $\vec{E}$		
Exams reflect what was taught	0.289*	0.025
Graded assignment returned promptly	0.069*	0.023
Graded assignment included comments	0.235*	0.024
Presented material effectively	1.280*	0.025
Was available for consultation	0.280*	0.026
Class rigor, $\vec{R}$		
Increased my ability to analyze and evaluate	0.191*	0.035
Course helped ability to solve problems	0.212*	0.033
Studied 1 or fewer hours per week	0.130	0.092
Studied 2 hours per week	0.212*	0.091
Studied 3 hours per week	0.271*	0.091
Studied 4–5 hours per week	0.218*	0.094
Studied 6–7 hours per week	0.174	0.109
Class grade-point average	0.135*	0.033
Grade student expected to earn	0.155*	0.018
Class externalities, $\vec{Z}$		
College standing, freshman	0.401	0.250
College standing, sophomore	0.270	0.249
College standing, junior	0.538*	0.233
College standing, senior	0.347	0.225
Class required for major	0.473	0.290
Class reported as required	0.331	0.376
Freshman in a required class	−0.333	0.314
Sophomore in a required class	−0.427	0.314
Junior in a required class	−0.764*	0.301
Senior in a required class	−0.608*	0.294
Freshman in a class reported as required	−0.193	0.406
Sophomore in a class reported as required	−0.178	0.403
Junior in a class reported as required	−0.551	0.386
Senior in a class reported as required	−0.455	0.379
Class enrollment	0.001	0.001
Class meets Monday, Wednesday, and Friday	−0.452	0.327
Class meets Tuesdays and Thursdays	−0.437	0.326
Class meets twice during the week	−0.414	0.329
Class meets once during the week	−0.378	0.329
Class meets between 8 and 10 a.m.	0.123	0.076
Class meets between 10 a.m. and 12 p.m.	0.187*	0.075
Class meets between 12 and 2 p.m.	0.178*	0.075
Class meets between 2 and 4 p.m.	0.211*	0.076
Class not located on the CA campus	0.211*	0.067
Modern, smart classroom	0.108*	0.041
Classroom not liked by faculty	0.057	0.034

Note: Model pseudo-R^2 = .439. The model correctly predicts 74.96% of 14,394 observations.
[a] A 95% confidence level is indicated by *.

Table 4. Marginal Effects[a]

Variable	Marginal Effects			
	Strongly Disagree	Disagree	Agree	Strongly Agree
Intercept	0.0755	0.2644	0.9575	−1.2974
Teaching effort, $\vec{E}$				
Exams reflect what was taught	−0.0036	−0.0127	−0.0459	0.0622
Graded assignment returned promptly	−0.0009	−0.0030	−0.0110	0.0149
Graded assignment included comments	−0.0029	−0.0103	−0.0373	0.0506
Presented material effectively	−0.0160	−0.0561	−0.2033	0.2754
Was available for consultation	−0.0035	−0.0123	−0.0445	0.0603
Class rigor, $\vec{R}$				
Increased my ability to analyze and evaluate	−0.0024	−0.0084	−0.0303	0.0411
Course helped ability to solve problems	−0.0027	−0.0093	−0.0336	0.0456
Studied 2 hours per week	0.0026	0.0090	0.0333	−0.0448
Studied 3 hours per week	0.0032	0.0112	0.0423	−0.0567
Studied 4–5 hours per week	0.0025	0.0089	0.0340	−0.0454
Class grade-point average	−0.0017	−0.0059	−0.0214	0.0290
Grade student expected to earn	−0.0019	−0.0068	−0.0247	0.0334
Class externalities, $\vec{Z}$				
College standing, Junior	0.0061	0.0215	0.0825	−0.1101
Junior in a required class	−0.0124	−0.0415	−0.1220	0.1759
Senior in a required class	−0.0089	−0.0305	−0.0974	0.1369
Class meets between 10 a.m. and 12 p.m.	0.0023	0.0080	0.0295	−0.0397
Class meets between 12 and 2 p.m.	0.0021	0.0075	0.0280	−0.0377
Class meets between 2 and 4 p.m.	0.0025	0.0087	0.0329	−0.0441
Class not located on the CA campus	0.0025	0.0087	0.0330	−0.0442
Modern, smart classroom	0.0014	0.0048	0.0173	−0.0235

[a] Variables associated with a 95% confidence level.

include only the intercept (−13749.32) reveals that the independent variables of the model explain 44% of the variation in SET scores, N (based on a pseudo R^2 = .439). Of particular interest is the predictive power of the model. This model predicted the correct outcome 75% of the time. By category, the model correctly predicted 107 of 376 responses of poor (28%), 337 of 1,061 responses of fair (32%), 3,370 of 4,322 responses of good (78%), and 7,093 of 8,635 responses of excellent (82%).

Twenty of 40 parameter estimates are statistically different from zero with a 95% confidence or better (Table 3). The marginal effects associated with these significant coefficients are reported in Table 4. As hypothesized, all the course effort, $\vec{E}$, variables improved the SET score, N, at the 99% confidence level. However, all the marginal effects except "presented material effectively" are relatively weak. Relative to the other five effort variables, "presented material effectively" has a far greater impact on SET scores, N.

In terms of course rigor, $\vec{R}$, seven of nine variables are significant at the 95% confidence level. Increasing a student's ability to analyze, evaluate, and solve problems increases the instructor's SET score. Apparently, students appreciate learning skills associated with evaluation and problem solving even if the class is difficult. Yet, as anticipated, if increased rigor and/or difficulty are associated with increased study time or a reduction in a student's expected class score and the class GPA, the likelihood is greater that the instructor's SET score will suffer. None of these rigor variables are particularly large relative to the others, and, in general, although influencing SET scores, N, are not a major influence (Table 4). Thus, rigor does appear to dampen an instructor's SET score, but well-presented rigorous material can easily offset this negative effect.

In terms of the externalities, $\vec{Z}$, these results support the hypothesis that factors beyond the control of faculty members influence their SET scores, N. At least for this data set, instructors who teach upper-division required courses in a student's major receive enhanced SET scores. Teaching predominantly juniors is detrimental to SET scores. Juniors required to complete some under-division courses may resent having to take these courses, which negatively impacts SET scores. Other significant but minor externalities on SET scores are class times, location, and smart classrooms. All these are negative externalities reducing SET scores. In particular, modern classrooms may improve instruction but tend to decrease SET scores. Given these significant externalities on SET scores, caution is called for in comparing SET scores not only across faculty but also in terms of individual faculty comparison across courses. Unless these externalities are internalized, SET scores are biased and should not be used for any type of evaluation.

Conclusions

In an effort to enhance their scholarly programs, faculty are faced with the constraint of allocating their limited resources toward scholarly activities and SET scores. For determining the optimal allocation of these resources, the individual influences resources have on SET scores is required. Based on economic theory, the research report in this article attempts to shed some light on the determinants of SET scores. Both determinants under the direct control of faculty and possible externalities affecting SET scores, which, by definition, faculty have no control over, are investigated.

The number one determinant for improving SET scores is the effort that faculty put into instruction. Specifically, instructors who take the effort to present course material effectively will have significantly improved SET scores. This type of effort dominates all the other determinants of SET scores. However, presenting material effectively does not necessarily mean using the latest electronic teaching tools. While the literature indicates classroom upgrades and transition to smart classrooms improves learning, the results of this investigation suggest it does not translate into improved SET scores. In fact, the results indicate just the opposite, the use of such tools actually decrease SET scores. Instructors allocating effort toward incorporating these tools into their classroom might want to reconsider their expected results. Further, college administrators may want to take a second look at the expected returns from smart classrooms. Such classrooms appear to be no substitute for effective course presentation. Converting lecture notes to PowerPoint presentations may not improve learning versus taking the time to write things on the chalkboard.

A result of this study, consistent with Baker, Rudd, and Hoover, indicates rigorous courses will lower SET scores. Faculty who derive satisfaction from offering rigorous and challenging courses must expend more effort toward improving presentations to offset the negative impact of rigor on SET scores. Alternatively, an instructor can just inflate his or her grades and improve SET scores. Caution is warranted in assessing effective instruction with SET scores, without also considering the grade distribution. High SET scores are a necessary but not a sufficient condition for quality instruction.

In contrast with effect and rigor, which faculty have direct control over, some major influences are not under their control. At the heart of this investigation are the impacts these externalities have on SET scores. The results indicate such externalities do have a significant impact on SET scores. This presents a possible bias in employing these scores as a measure of instructor effectiveness. Class times, location, and smart classrooms can lower an instructor's SET scores.

These effects of rigor and externalities on SET scores support the general theory emerging from the literature that quality of instruction cannot be compared across faculty or courses with SET scores. In fact, this theory has been upheld so extensively that it should be confirmed as a paradigm in the economics of education literature. The results from this research extend the literature by empirically demonstrating the biases in SET scores introduced by externalities. Using SET scores to evaluate even the same instructor teaching the same course over time may be invalid. In response to this paradigm of biased SET scores, universities are slowly retreating from SET scores as the sole criterion for teaching effectiveness. In its place are teaching portfolios, which provide a more comprehensive assessment of an instructor's effectiveness.

[Received June 2003; Accepted May 2005.]

References

Baker, M., R. Rudd, and T. Hoover. "Relationships Between Student/Course Characteristics and Student Evaluations of Teaching Quality." *North American Colleges and Teachers of Agriculture Journal* 44(2000):25–29.

Becker, W.E., and M. Watts. "How Departments of Economics Evaluate Teaching." *American Economic Review* 89(May 1999):344–49.

Braskamp, L., E. Brandenburg, E. Kohen, J. Ory, and P. Mayberry. "Guidebook for Evaluating Teaching, Part I." *North American Colleges and Teachers of Agriculture Journal* 27(December 1983):29–33.

———. "Guidebook For Evaluating Teaching, Part II." *North American Colleges and Teachers of Agriculture Journal* 28(March 1984):19–24.

Frey, P.W. "Validity of Student Instructional Ratings: Does Timing Matter?" *Journal of Higher Education* 47(1976):327–36.

Greene, W.H. *Econometric Analysis*. New York: Macmillan Publishing Company, 1990.

Greenwald, A.G. "Validity Concerns and Usefulness of Student Ratings of Instruction." *American Psychologist* 52(1997):1182–86.

Greenwald, A.G., and G.M. Gillmore. "Grading Leniency Is a Removable Contaminant of Student Ratings." *American Psychologist* 52(1997):209–217.

Seevers, B.S., T.J. Dormody, and D.M. VanLeeuwen. "Developing a Valid and Salable Student Evaluation of Teaching (SET) Instrument." *North American Colleges and Teachers of Agriculture Journal* 43(December 1999):15–19.

Wilson, R. "New Research Casts Doubts on Value of Student Evaluations of Professors." *The Chronicle of Higher Education* 54(1998):12–16.

Worley, T., and K. Casavant. "Student Evaluations of Teaching: A Tool for Directing and Measuring Course Improvement Efforts." *North American Colleges and Teachers of Agriculture Journal* 39(1995):37–38.