Pitfalls in Evaluating Instruction: The Case of Higher Education

September 2007

Bruce A. Weinberg
The Ohio State University, IZA, and NBER
weinberg.27@osu.edu

Belton M. Fleisher
The Ohio State University and IZA
fleisher@econ.ohio-state.edu

Masanori Hashimoto
The Ohio State University
mhashi@econ.ohio-state.edu

Abstract

This paper studies methods for evaluating instruction in higher education. We explore student evaluations of instruction and a variety of alternatives. We develop a simple model to illustrate the biases inherent in student evaluations. Measuring learning using grades in future courses, we show that student evaluations are positively related to current grades but uncorrelated with learning once current grades are controlled. We offer evidence that the weak relationship between learning and student evaluations arises in part because students are not aware of how much they have learned in a course. We conclude with a discussion of alternative methods for evaluating teaching

We are grateful for comments from Tisha Emerson, Eric Fisher, and Hajime Miyazaki, seminar participants at Ohio State University and especially members of the Ohio State University Undergraduate Economics Society, and participants at the 2007 American Economic Association Meetings. We are also grateful for able research assistance by Xueyu Cheng and Young-Kyu Moh and for assistance with data assembly by John-David Slaughter.

I. Introduction

This paper considers methods for assessing performance in higher education. Evaluations are important for diagnosing and correcting teaching problems and for faculty performance reviews. Moreover, with the Miller Commission's focus on accountability in higher education, evaluation methods are receiving increasing public attention (United States Department of Education [2006]; Golden [2006]). While assessment is never easy, the wide-range of subjects taught makes assessment in higher education particularly difficult. Given the costs and difficulty of comprehensive assessments of teaching effectiveness, student evaluations traditionally have been the primary, if not the only, means of assessing teaching in higher education.[1]

We employ three constructs to assess teaching performance. First, we employ student evaluations of teaching. We also estimate learning in a section based on the average grade that the students in a section receive in subsequent classes. An advantage of this measure is that it weights knowledge in proportion to its importance in future work. Lastly, we study the factors that determine whether students take additional courses in the area.[2] The second and third measures have the advantage of being intuitively simple and only require information that is readily available.[3] We study the relationship between these performance measures and how instructor characteristics including gender, foreign birth, tenure-track status, and graduate-student status, as well as course characteristics relate them.

Given the prominence of student evaluations in measuring teaching effectiveness, much of our analysis focuses on them, beginning with a simple model of the determinants of student evaluations. We use our model to clarify the factors that affect evaluations, to study the welfare implications of using student evaluations to evaluate teaching, and to develop alternative evaluation criteria.

---

[1] White (1995) reports that SEI scores are the predominant measure of teaching effectiveness used by economics departments. He notes that there appears to be strong reluctance to rely on direct observation of teaching, particularly among research-oriented departments.

[2] Two obvious additional measures are drop rates and wait lists. Wait lists are uncommon in these courses and the measure of drop rates includes students who dropped before the beginning of the class as well as those who dropped once the course began.

Both economists and educational psychologists have studied how grades and learning affect student evaluations but, to the best of our knowledge, no studies consider both factors together. We assume that students care about their grades, the amount they learn, and the course experience. We show that in order to estimate the effects on student evaluations of learning or human capital and "grading leniency" – grades conditional on learning – both variables must be included simultaneously. This is because classes in which students learn more may receive higher grades and thus give higher evaluations. We also show that student evaluations likely place too little weight on learning and generate an incentive for instructors to inflate grades.

The present work fits into an emerging literature on the determinants of outcomes in higher education (see Bettinger and Long [2004]; Beddard and Kuhn [2005]; and Hoffmann and Oreopoulos [2006]). It relates to a large literature in the economics of education on the determinants of student outcomes in primary and secondary education.

There is also a rich literature on student evaluations of teaching, comprising thousands of pieces (Feldman [1997]), in educational psychology. Fortunately there are a number of large-scale literature reviews. This literature focuses on the relationship between student evaluations and various measures of learning and generally finds a strong relationship. The most persuasive evidence for a link between learning and evaluations comes from multi-section courses with common syllabi and exams (Cohen [1981]; Dowell and Neal [1982]; Marsh [1984, 1987, 2006]; Abrami, d'Apollonia, and Rosenfield [1997]; Feldman [1997]; and Theall and Feldman [2006]). The lack of instructor discretion in these courses leads one to question the extent to which these results will generalize to other courses. More importantly, this design is not suitable for separating the effect of grades from that of learning on evaluations because there is little if any variation in grades conditional on learning. Discussions with students suggest that they estimate how much they have learned in a course from the grade that they expect to receive. If so, in multi-section classes, students' estimates of their learning will be highly correlated with their grades. The literature has noted that estimates of the effect of learning on student evaluations

---

[3] Alternatives would include drop rates and the length of wait lists, neither of which are available in our data.

may be biased by grades. Because the multi-section design essentially makes it impossible to separately estimate the effect of grades and learning, educational psychologists have generally relied on indirect methods to address the effect of grading leniency (see, for instance, Greenwald and Gillmore [1997]).[4]

The economics of education literature has provided evidence that student evaluations are related to grades and argued that the use of evaluations may lead to grade inflation.[5] Empirical work linking expected grades to evaluations is mixed.[6] We depart from this work in two ways. First, we use actual course grades rather than expected grades. While students generally do not receive grades until after completing their evaluations, students have some idea of what grades they may receive based on midterm results; homework scores; and other objective information on their course performance as well as conscious or unconscious indications from the instructor. Second, unlike most of the literature, we measure grades using the average grade in a section rather than at the student level. In an individual-level regression, most of the variation in grades arises from individual differences in grades within a section. So the individual level relationship between grades and evaluations indicates whether students who are at the top of a given section give higher evaluations than those at the bottom of that section, not whether instructors who grade more leniently receive higher evaluations.

While educational psychologists and economists have both studied how student

---

[4] In the economics literature Sheets, Topping, and Hoftyzer [1995] employ a multi-section approach. Shmanske [1988] uses grades in a subsequent course in a two-course sequence, which is related to our approach but much less widely applicable. Neither study includes current grades.

[5] Becker and Watts (1999) criticize economics departments for "following the herd" in their uncritical use of SEI measures and not applying the same rigor they require of published research to the use and understanding of "teaching-quality" survey instruments to evaluate the performance of their faculty. Kanagaretnam, Mathieu, and Thevaranjan (2003) cite several articles from the *Chronicle of Higher Education* dealing with the topic of the impact of SEI's on student learning and grade inflation. McKenzie (1975) develops a simple model of consumer choice in which the use of SEI's by academic institutions provide an incentive for instructors to alter the grade-effort tradeoff that students face to make it easier (less costly in terms of effort) to earn higher grades. This contributes to grade inflation and adversely affects the institution's ability to distinguish good and bad students. A search of the *Chronicle's* table of contents for the key words "student evaluation" yields 22 articles and notes for the year 2005 (through the end of October). A search for both "student evaluation" and "grade inflation" yields six letters and articles between 1998 and 2005, for example, Benton (2004).

[6] Nichols and Soper [1972]; Krautmann and Sander [1999]; Boex [2000]; and Kelley [1971] report a positive relationship between expected grades and evaluations, while DeCanio [1986] and Nelson and Lynch [1984]; and

evaluations are related to course grades or learning, to the best of our knowledge, this paper is the first to study how grades and learning are jointly related to evaluations.

Our data cover nearly fifty thousand enrollments in almost four hundred offerings of principles of microeconomics, principles of macroeconomics, and intermediate microeconomics over a decade at The Ohio State University.[7] We find a strong positive relationship between student evaluations and both current and future grades when they are included separately, but when they are included in the same model, the current grade is related to student evaluations but future-course grades are not. There are many potential explanations for these results, including a variety of selection arguments. We devote considerable effort to six of them, concluding that, on average, students are not aware of the amount of human capital produced in a class. We have no reason to believe that the focus on current grades and uncertainty about learning is specific to economics or the institution studied and therefore expect our results to generalize at least qualitatively.

As indicated, we also study the number of subsequent economics classes that students take as a revealed-preference measure of quality (see Hoffmann and Oreopoulos [2006]). Further eroding our confidence in student evaluations, we find that they are unrelated to the number of subsequent economics classes that students take.

We consider how instructor characteristics are related to our measures of the quality of instruction. In some cases, female and foreign-born instructors receive lower student evaluations than male and US born instructors. Learning, however, is not related to instructor gender or national origin, nor do we find systematic differences in evaluations or student learning between non-tenure track faculty and tenure track faculty. This finding is noteworthy in light of Ehrenberg's (2004) observation that we have little or no knowledge of the effect of part-time and non tenure-track faculty on student learning and other measures of academic production. While

---

Bosshardt and Watts [2001] report weak, negative, or mixed results.

[7] These courses were chosen because they are standard, they enroll the most students, and more of the students in these classes take additional economics classes. These were the only classes for which data were collected or analyzed.

we do not find that observable instructor characteristics are related to learning, we do find large variations across instructors in performance. This result is consistent with evidence from primary and secondary education (See Rivkin, Hanushek, and Kain [2005]).

III. Analytical Framework

Because student evaluations of teaching are the most common method of evaluating instruction, we begin with a simple model of student evaluations. Our model is intended to clarify what student evaluations measure and illustrate the potential biases in estimating the determinants of evaluations. We are also interested in the extent to which rewarding instructors based on student evaluations yields desirable policy outcomes.

We focus on a representative (male) student and highlight three factors, grades, human capital, and the course experience. Let h_s denote human capital produced in the course, where s indexes sections. Grades, $g_s \in \lfloor \underline{g}, \bar{g} \rfloor$, which are discussed at greater length below, are a function, $G(h_s)$ (where $G' \geq 0$), of human capital acquired in the course and the leniency of the (female) instructor in assigning grades, δ_s. Formally, $g_s = G(h_s) + \delta_s$. When completing evaluations, the representative student will form expectations of his grades based on feedback from the instructor, but will neither know his performance on the final examination nor will he have complete information about any curve.[8] Let $E[g_s] \in [\underline{g}, \bar{g}]$ denote the representative student's expectation of his grade at the time of the evaluations. The representative student's course experience, x_s, represents the (dis-) utility derived from the course, including the disutility of coursework.

We assume that students give a scalar evaluation, e_s, determined by the evaluation function given by:

$$e_s = \phi(E[g_s], h_s, x_s). \tag{1}$$

We think of this as the student's utility function, but this assumption is not necessary for much of

[8] In our empirical work, we proxy for the expected grades, which is not observable, by using actual grades and grades in previous offerings of the course by the instructor. These can be thought of as reflecting rational and

the analysis. Presumably, evaluations are non-decreasing in expected grades and human capital produced in the class ($\phi_g \geq 0$ and $\phi_h \geq 0$). We normalize x_s to be a good so that increases in the course experience increase evaluations ($\phi_x \geq 0$).

As our literature review makes clear, past studies have found that when included separately both learning and grades are positively related to evaluations. The positive relationship between learning and evaluations has been taken as evidence that evaluations reflect learning, while the positive association with grades has been taken as evidence that grading leniency increases evaluations. To assess these claims, we return to the evaluation function, $e_s = \phi(E[g_s], h_s, x_s)$. It is convenient to assume that this function is linear,

$$e_s = \phi_g g_s + \phi_h h_s + \xi_s . \tag{2}$$

The error, $\xi_s (= x_s)$, represents the course experience. Ignoring the possibility that the error is correlated with learning or grades (at least for the moment), either interpretation of the relationships between evaluations and learning or between evaluations and grades could be wrong. If instructors give higher grades when students learn more, i.e. if $G' > 0$ then higher evaluations in courses with more learning may simply indicate that students like high grades, i.e. $\phi_g > 0$, even if the do not care about learning, i.e. $\phi_h = 0$. Alternatively, the positive relationship between grades and evaluations may indicate that instructors give higher grades when students learn more, i.e. $G' > 0$, and that students reward their instructors for added human capital, i.e. if $\phi_h > 0$. In such a case, evaluations will be associated with higher grades even if grades do not directly affect evaluations, i.e., even if $\phi_g = 0$.[9]

Educational psychologists have argued that if there is little effect of grades on evaluations, but a strong relationship between grades and learning, one would not want to adjust student evaluations for grades (Greenwood and Gillmore [1997]). It is clear, however, that if evaluations are affected by grades then a policy maker would want to adjust evaluations for

---

adaptive expectations respectively.

grades, provided that one can condition on learning. In addition to removing a source of noise in evaluations, doing so would reduce instructor incentives to inflate grades.

To further analyze the problem of interpreting student evaluations, we assume that the instructor has a technology governing the relationship between the course experience and learning. We write the production possibility frontier for the instructor of section *s* as $0 = \pi(x_s, h_s; \theta_s)$. Here θ_s denotes an idiosyncratic instructor effect (which might be thought of as "ability"). We begin by assuming that θ_s is exogenous, but relax this assumption later. We assume that, for a given value of θ_s, greater human capital production is associated with a less pleasant course experience for students because more work effort is required of them $(\left.\frac{dx_s}{dh_s}\right|_{\theta_s} \leq 0)$.[10] To simplify the analysis, we assume that the marginal rate of transformation between the course experience and human capital is declining – when most class time is spent on producing learning, engaging material can be introduced with relatively little cost to learning, but as more and more time is spent on entertainment, learning suffers more. Figure 1 illustrates a production possibility frontier.

The student's iso-evaluation curve, $e = \phi(E[g_s], h_s, x_s)$, is assumed to have a declining marginal rate of substitution between the course experience and learning. Figure 1 also shows iso-evaluation curves for a given level of expected grades, $E[g]$. The tangency point indicates the (h_s, x_s)-pair that maximizes evaluations, given θ_s.

We assume the instructor maximizes her utility, which depends on evaluations, the human capital she produces, and the grades she gives, so that $U(e_s, h_s, g_s)$, where $U_e \geq 0$. The instructor's utility depends on human capital directly through *h*, and indirectly through evaluations, *e*. The instructor's utility also depends indirectly on the course experience (*x*)

---

[9] To simplify the discussion, we ignore the course experience here.

[10] If improvements in the course experience increase student engagement, the production possibilities frontiers might slope up in some regions, but an instructor will always locate in a region where there is a negative

through *e*. We introduce grades into the instructor's utility function under the assumption that she derives disutility from giving grades that are inconsistent with student performance and the norms at the institution and may incur costs imposed by administrators if grades deviate significantly from institutional norms. We assume that that $g \in \left[\underline{g}, \bar{g}\right]$ and that *U* is strictly convex and hump-shaped in *g* with $U_g\left(e_s, h_s, \underline{g}\right) > 0$ and $U_g\left(e_s, h_s, \bar{g}\right) < 0$, so that utility is increasing in grades near their minimum value, $\underline{g}$, and decreasing in grades near their maximum $\bar{g}$. We also assume that $U_{gh} \geq 0$ so that the marginal utility of increasing grades is higher (or less negative) for better students.

By incorporating the function governing student evaluations into the instructor's utility function, we can write

$$\tilde{U}\left(e_s, h_s, g_s\right) = U\left(\phi\left(g_s, h_s, x_s\right), h_s, g_s\right).$$

This function is concave in h_s and x_s under our assumptions.

The instructor chooses both human capital and grades, while students choose how to evaluate her. The first order conditions for a maximum to the instructor's problem are

$$\frac{d\tilde{U}}{dg_s} = U_e \phi_g + U_g = 0$$

$$\frac{d\tilde{U}}{dh_s} = U_e \frac{de_s}{dh_s} + U_h = 0 \quad where \quad \frac{de_s}{dh_s} = \phi_h - \phi_x \frac{\pi_h}{\pi_x}.$$

The first order condition for the grade says that as long as the instructor cares about evaluations and so long as higher grades lead to better evaluations, then she sets grades in the region where she receives disutility from raising grades further. At her optimum, the marginal disutility of raising grades equals the marginal utility from the increase in evaluations induced by higher grades. Variations in the marginal disutility of raising grades generate variations in grading leniency and evaluations. The first order condition for human capital says that the instructor sets *h* where the marginal utility of higher *h* equals the reduction in utility from lower evaluations.

---

relationship between learning and the course experience.

We begin by analyzing the special case where human capital does not directly enter the instructor's utility function so that she sets *h* to maximize her evaluations. This case is particularly easy to characterize and provides a convenient benchmark for the more general case. In this case $U_h = 0$, so that the instructor sets $\frac{de_s}{dh_s} = U_e\left(\phi_h - \phi_x \frac{\pi_h}{\pi_x}\right) = 0$, implying that $\frac{de_s}{dh_s} = \phi_h - \phi_x \frac{\pi_h}{\pi_x} = 0$. This point is shown in Figure 1 as point *a*, where the instructor's production possibility frontier is tangent to the representative student's indifference curve.

If the instructor cares about human capital production directly, she will deviate from the student's optimal level of human capital. It is plausible that instructors sometimes choose content that is more difficult than the level that maximizes evaluations, either because they perceive harder content as being consistent with their duties as instructors (see below for a justification) or because they derive more utility from more rigorous content. When the instructor receives utility directly from human capital production, $U_h > 0$ she is willing to incur a cost in the form of lowered evaluations so as to produce more human capital. In figure 1, this corresponds to a movement along an instructor's production possibility frontier to the tangency at (*b*) between the production possibility frontier and the dotted indifference curve for the instructor. As shown, evaluations are lower at this point than at the one that maximizes evaluations.[11] Thus, variations in the taste for challenging content causes a downward bias in the relationship between human capital and evaluations (for a given technology), because a professor who exerts more effort producing human capital provides a worse (unmeasured) course

[11] To see this formally, note that a change in the instructor's utility function that raises U_h leads the instructor to produce more human capital. That is, $\frac{dh}{dU_h} = -\frac{1}{U_{ee}\left(\frac{de}{dh}\right)^2 + U_e \frac{d^2e}{dh^2} + U_{hh}} > 0$. To the right of (a), $\frac{de}{dh} = \phi_h - \phi_x \frac{\pi_h}{\pi_x} < 0$, so $\frac{de}{dU_h} = \frac{de}{dh}\frac{dh}{dU_h} < 0$.

experience.[12]

As indicated above, estimates of the evaluation function (2) that do not include both grades and human capital are likely to be biased. Moreover, even if measures of both grades and human capital are available, there is a problem with estimating the evaluation function (2). In particular, in order to obtain unbiased estimates of ϕ_g and ϕ_h in (2), the course experience ξ_s must be uncorrelated with grades and human capital. Our model implies that learning may well be correlated with the course experience. An outward shift in the professor's production possibilities frontier will likely lead to higher levels of both human capital and the course experience, which would bias upward the estimate of the coefficient on learning.

Thus far we have assumed that the idiosyncratic instructor effect, or her "ability," θ_s, is exogenous. However, instructors may be able to improve their teaching by exerting more effort. To capture this possibility, we can augment the instructor's utility function to allow for a cost of increasing θ_s,

$$\tilde{\tilde{U}}(e_s, h_s, g_s, \theta_s) = U(F(g_s, h_s, x_s), h_s, g_s) - c(\theta_s).$$

In this case, in addition to the first order conditions for the instructor's problem above, we have the condition that

$$\frac{d\tilde{\tilde{U}}}{d\theta_s} \Rightarrow \left(U_e \frac{de_s}{dh_s} + U_h \right) \frac{\pi_\theta}{\pi_h} = c'(\theta_s) = U_e \frac{de_s}{dx_s} \frac{\pi_\theta}{\pi_x}.$$

Thus, the instructor sets the marginal cost of effort equal to its marginal benefit as measured by the change in utility from additional human capital production and/or a better course experience. Otherwise the implications are unchanged.

Optimal Evaluation Criteria

The preceding analysis points to a number of concerns when relying on student

[12] An instructor who obtained disutility from producing human capital might locate at a point like (*c*). Here too, evaluations are lower than at (*a*), but in this case a positive correlation between evaluations and learning emerges.

evaluations to assess teaching effectiveness. First, to determine how student evaluations are affected by grades and learning, one must control for both variables simultaneously. Second, the effect of human capital on evaluations may be biased by a correlation between grades and the unmeasured course experience.

Student evaluations also depend on students' assessment and valuation of how much they have learned in a course. Below, we provide evidence that they may be not be well positioned to make that determination. Even if students accurately assess their learning, they may place less weight than institutions on learning relative to the course experience. For example, society and parents may place higher weight on human capital production and less weight on the course experience than students do because students discount at a high rate, or because human capital generates externalities for society. In either case, relying solely on student evaluations can distort instructors' incentives away from the social optimum.

To address these issues, we consider a social planner who places weight λ_s on the student, weight λ_I on the instructor, and derives benefit $\Lambda(h_s)$ from human capital (because of external effects or as a response to excessive discounting by students). The social welfare function is

$$V(g_s h_s, x_s) = \lambda_S \phi(g_s, h_s, x_s) + \lambda_I U(e_s, h_s, g_s) + \Lambda(h_s).$$

The marginal rate of substitution between the course experience and human capital for a social planner who can adjust the evaluation criterion to neutralize any effect of evaluations on the instructor's utility, is

$$MRS = \frac{\frac{\partial V}{\partial h_s}}{\frac{\partial V}{\partial x_s}} = \frac{\lambda_S \phi_h + \lambda_I U_h + \Lambda'}{\lambda_S \phi_x}.$$

Thus, the social planner's indifference curves will be steeper than the student's indifference

curves (given by $\frac{\phi_h}{\phi_x}$), but may be more or less steep than the instructor's indifference curves.

If a direct measure of human capital and unbiased estimates of $\hat{\phi}_g$ and $\hat{\phi}_h$ are available, it is possible to estimate the course experience directly. Instructors can be evaluated on the course experience they provide, the grades they assigned, and the amount of human capital they produced. With unbiased estimates of $\hat{\phi}_g$ and $\hat{\phi}_h$, the course experience can be estimated by

$$\hat{x}_s = e_s - \left(\hat{f}_g g_s + \hat{f}_h h_s\right). \tag{3}$$

With a sense of social priorities, estimates of the course experience, human capital, and grades, administrators can reward instructors based on social welfare.

IV. Data

Our data set includes students who took principles of microeconomics, principles of macroeconomics or intermediate microeconomics at The Ohio State University between 1995 and 2004. We obtained data on all subsequent economics courses taken by these students through the end of the 2004 academic year. The data set includes identifiers for the sections the students took, student demographic characteristics, and grades in all economics courses taken during this period. Student evaluations are anonymous and are available at the section level but not at the student level. Thus, we estimate the relationship between grades and evaluations at the section level rather than at the individual level, which we believe is correct for the reasons discussed above.

Our evaluation instruments contain 10 items, including an overall score, which is our primary focus. The other questions are shown in the tables and include measures of perceived learning, preparation and organization, the instructor's attitude, and the extent to which the course stimulated students to think. Table 1 contains the variable definitions and their means and standard deviations for the three sets of courses. Our data set comprises 194 sections (with 28,172 students) in principles of microeconomics; 122 sections (with 15,809 students) in principles of macroeconomics; and 88 sections (with 4,428 students) in intermediate

microeconomics. The average evaluation score ranges from 3.72 (standard deviation of .54) for principles of macroeconomics to 3.86 (standard deviation of .44) for principles of microeconomics on a scale of 1 (lowest) to 5 (highest). On a four-point scale, the average course grade is close to 2.7 (with a standard deviation of about .3), a B-, for all three courses. The table shows the distribution of instructor and student characteristics for the three courses.

V. Estimation

We employ a multi-step strategy to estimate grades and learning and their relationship to student evaluations. We first estimate grades. Then we estimate the amount of learning in each section based on grades in subsequent sections. These learning estimates are of interest in their own right and we study them and also use them to estimate how grades and learning are related to evaluations. This section describes our procedure step-by-step in terms of principles of microeconomics, including how we merge our individual-level data on current and subsequent grades into the section-level data to be compatible with our section-level evaluations. Our procedures for principles of macroeconomics and intermediate macroeconomics are similar.

Step 1. Estimating Grades

Let *i* index students and *s* index the base section (*i.e.* the particular section of principles of microeconomics that the student took). Let g_{is} denote the grade received by student *i* who took base section *s*. In the first step, we regress g_{is} on a vector of base section dummy variables $\vec{D}_{is}$ and, in some specifications, the student's characteristics at the time of the base section, $\vec{X}_{is}$. Our specification is

$$g_{is} = \vec{X}'_{is}\vec{\beta}_1 + \vec{D}'_{is}\vec{\psi} + \varepsilon_{1is} \qquad (*)$$

The coefficient ψ_s on the dummy variable for base section *s* gives the mean grade in the section (with or without controls for individual characteristics). These coefficients are used in our third stage to capture grades.

Step 2. Estimating Learning

To estimate learning, we use grades in subsequent courses. Let *j* index sections of

subsequent economics courses, so that g_{isj} denotes the grade of student *i*, who took base section *s*, in subsequent section *j*. We regress the grades in subsequent courses, g_{isj}, on a vector of dummy variables for the subsequent section (to control for differences in grading across classes), $\vec{Z}_{isj}$; a vector of dummy variables for the base section, $\vec{D}_{isj}$; and, in some specifications, student characteristics, $\vec{X}_{isj}$, at the time of section *j*. Formally,

$$g_{isj} = \vec{X}'_{isj}\vec{\beta}_2 + \vec{Z}'_{isj}\vec{\Gamma} + \vec{D}'_{isj}\vec{\theta} + \varepsilon_{2isj} \qquad (**)$$

The coefficient θ_s on the dummy variable for students who took base section *s* indicates how well these students do in later courses. This coefficient is our measure of the learning that took place in section *s*. The set of controls can be varied to include measures of student ability so that θ_s reflects learning, or these variables can be excluded so that the θ_s indicates human capital at the end of the course. These estimates are of interest in their own right and are used in our third step to control for human capital in the section.

Step 3. Evaluating Student Evaluations

In the first step, we estimated ψ_s, the grades in base section *s*; in the second step we estimated θ_s, the learning in base section *s*. In the last step, we regress the student evaluations for base section *s*, e_s, on learning, grades, and instructor and section characteristics, $\vec{W}_s$:

$$e_s = \theta_s\phi_h + \psi_s\phi_g + \vec{W}'_s\vec{\rho} + \varepsilon_{3s}. \qquad (***)$$

The coefficient ϕ_h tells us how much students value learning (net of any costs of learning) and the coefficient ϕ_g, how much students value high grades when evaluating the instructor. The coefficient vector $\vec{\rho}$ tells how observable instructor characteristics are associated with evaluations.

We spend considerable time addressing alternative explanations of our results, including selection issues. Our methods employ a variety of strategies, as will be discussed later.

Additional Analyses

We can also estimate a variety of related effects. We estimate the effect of instructor characteristics such as gender, native language, tenure track status, or whether or not the instructor is a graduate teaching associate, on learning. To do this, we estimate,

$$\theta_s = \vec{W}_s' \vec{\beta}_3 + u_s . \qquad (****.1)$$

As above, W_s would represent the characteristics of section *s*, including those of the instructor.

We also assess how instructor characteristics are associated with grading leniency, by estimating,

$$\psi_s = \vec{W}_s' \vec{\beta}_4 + \theta_s \gamma + \xi_s . \qquad (****.2)$$

One could estimate this model with or without θ_s as a control for the effect of human capital.

VI. Findings

This section reports our estimation results for equation (***). We begin with our main results for principles of microeconomics, and then discuss the results for principles of macroeconomics and intermediate microeconomics. We then turn to alternative explanations of our results, including those based on selection issues, and conclude with some additional analyses.

Principles of Microeconomics

The first column of table 2 reports a regression of student evaluations on the current course grade. We find that students in sections with higher grades rate their courses more highly than those in other sections. Column 2 reports a regression with only our learning measure, i.e., future grades, which are also found to be positively associated with evaluations, but with a smaller coefficient. When both current and future course grades are included in the same regression (column 3), the effect of current grade clearly dominates, and the coefficient for future grade is small and insignificant.

The remaining columns examine a variety of other potential determinants of student evaluations. First, we include a set of instructor characteristics without controlling for the grades

(column 4). Female instructors receive lower evaluations than men, as do foreign-born instructors, although this latter difference is not statistically significant. (Half as many sections are taught by foreign instructors as by women, making this estimate imprecise.) There are no discernable differences in evaluations between non-tenure track lecturers, graduate teaching associates, and tenure-track faculty.

Differences in grading practices and learning may be responsible for the gender gap in evaluations as well as the substantial (but statistically insignificant) foreign-domestic gap. To explore this possibility, we include both current and future course grades along with instructor characteristics, in column (5) of the table. The inclusion of these variables does little to the gender and foreign-domestic gaps in evaluations. The above evidence suggests that students rate women and perhaps foreign instructors less favorably than others, possibly reflecting distaste/disrespect for such instructors or unmeasured differences in the course experience like language ability or teaching style.

The regressions in columns (6) and (7) include characteristics of the students in the course and then section characteristics; column (8) reports estimates with all of these variables, year dummy variables, and the response rate for the evaluations in the section. In both regressions, the coefficient for current course grade are significant and similar in magnitude. To summarize other statistically significant findings, column (6) shows that sections with more black students rate their instructors more highly. Column (7) shows that students in night classes give statistically significantly higher evaluations than other classes. Column (8) shows that the coefficients for blacks and night classes as well as for female instructors all are significant after fully controlling for the available variables. The foreign effect remains large, but insignificant.

The estimates in Table 2 consistently show a statistically significant effect of the current course grade. Indeed, the coefficient becomes larger as more variables are controlled. According to column (8), a one standard-deviation change in the current course grade is associated with a large increase in evaluations – over a quarter of the standard deviation in evaluations. Once current grades are controlled, learning, as measured by future grades, is never

statistically significantly related to evaluations.

Our use of the actual current course grade as a measure of expected grade in the course deserves some discussion. As indicated above students likely have some idea of what grades they will receive based on formal or informal feedback received during the quarter. Alternatively, students may form expectations of their course grade based on the reputation the instructor's grading in previous offerings of the course. We examine this last possibility by including in our regression the lagged grade – the mean grade in the last offering of the course by the instructor – along with the current grade. Column (9) presents results without the lagged grade for the sample for which the lagged grade is available. Including lagged grade, in column (10) does not change the estimated coefficient of the current course grade or the future grade, and the coefficient for the lagged grade is itself small and statistically insignificant. It appears that students base their evaluations on indications provided by the professor about the current course rather than on the professor's reputation (at least based on recent offerings of the course).

Individual Evaluation Items

The evaluations we use have ten items, nine focusing on specific aspects of the course experience and an overall score, which has been the focus of the analysis thus far. We now turn to the individual items. The various items are highly correlated, with none of the correlations beneath .75 and most above .8 or .9.[13]

Table 3 reports estimates with the individual evaluation items as the dependent variables. The estimates for these individual items are quite similar to those for the overall evaluation measure. The current course grade is always associated with higher evaluations and the relationship is statistically significant at the 5% level in 9 of the 10 cases. None of the evaluation items is statistically significantly related to future grades. Assuming that future grades reflect learning in the current course, these findings suggest that grading leniency, but not learning, has a significant impact on student evaluations. While not always statistically significant, women

[13] Sarwark *et al.* (1995) point out that instructor "halo" effects may affect all items.

and foreign instructors tend to receive lower evaluations, black students tend to give higher evaluations, and honors and night classes tend to give higher evaluations.

Evaluations on organization and preparation have the weakest relationship with current grades. It is also noteworthy that the item that captures learning, "Learned greatly from instructor" is no more closely related to future grades than any of the other items. This finding suggests that students are not able to evaluate the amount they learn in a course or that they base their estimates on the grades that the expect to receive. Alternatively, the later course grade may not be a good proxy for learning in the current course perhaps because what is learned in the current course has little bearing on later course, a possibility we consider below.

Principles of Macroeconomics

This section reports results for principles of macroeconomics. These results are presented in the same order as those for principles of microeconomics and are generally consistent with those for principles of microeconomics. We note that there are only 60% as many macro-principles sections as there are micro-principles sections and that fewer of the students in macro-principles take subsequent classes, so the estimates for macro-principles are somewhat less precise than those for micro-principles.[14]

The estimates in the top panel of Appendix Table 1 show that grades in the current course are strongly related to student evaluations for later courses. In fact, the estimates are slightly larger than those for micro-principles. Grades in future courses are unrelated to evaluations, whether they are included on their own or with the current course grade. Again women and foreign born instructors tend to receive lower evaluations than men and domestic instructors, but these differences are not systematically statistically significant.

The top panel of Appendix Table 2 reports estimates for each of the individual survey items. These estimates show a positive relationship between grades in the macro-principles

[14] While micro-principles is not a prerequisite for macro-principles, almost all students take micro-principles before macro-principles, so that almost all of the grades in the macro-principles classes are available as subsequent grades for the micro-principles estimates, while for most students in macro-principles subsequent grades are only available for students who take a third economics course.

section and grades in subsequent courses for all ten survey items (the relationships are statistically significant at the 5% level in 7 of the 10 cases). There is no evidence of a positive relationship between grades in later courses and any of the evaluation items. As with micro-principles, the weak relationship between grades and later courses holds true for the item "learned greatly." As above, there is some tendency for foreign instructors to receive lower evaluations.

Intermediate Microeconomics

This section presents results for intermediate microeconomics. Again, there are fewer intermediate economics sections than micro-principles (under half as many) or macro-principles (three quarters as many) and fewer students take later classes making the estimates noisier.[15]

Bearing this caveat in mind, the estimates in the bottom panel of Appendix Table 1 consistently show a positive relationship between current course grades and evaluations. Some of the intermediate microeconomics sections use calculus. Calculus based sections receive higher evaluations, the students in these sections tend to receive higher grades, and they tend to receive higher grades in future courses (both of these results are in Table 7 and are discussed below). Controlling for whether the course was calculus-based increases the relationship between current grades and evaluations and generates a negative (but insignificant) relationship between later grades and evaluations.

The bottom panel of Appendix Table 2 presents results for the individual survey items. Current course grades are positively related to evaluations while later course grades are negatively related to evaluations for all of the individual items. For 6 of the 10 items the current course grade is statistically significantly positive at the 10% level, while later course grades are statistically significantly negative for 3 of the 10. Calculus-based sections tend to give higher evaluations, as do night sections. There is some tendency for instructors with Ph.D.s to receive lower evaluations.

[15] Many business majors require intermediate microeconomics, but no additional classes.

Summary

The highlights of what we found so far are:

1. We consistently find a positive relationship between grades in the current course and evaluations. This finding is robust to the inclusion of a wide range of controls.
2. There is no evidence of a positive relationship between learning and evaluations controlling for current course grades.
3. Learning is no more related to student evaluations of the amount learned in the course than it is to student evaluations of other aspects of the course.
4. In some cases women and foreign-born instructors receive lower evaluations than other instructors, all else equal.

VII. Do Future Grades Capture Learning?

On the assumption that our measure of learning as valid, the preceding findings imply that grading leniency is an important determinant of evaluations and that students do not reward instructors who generate learning per se. We offer six alternative explanations for these findings.

First, they may indicate that grades in future courses are noisy measures of learning. Second, they may indicate that there is selection into courses – for instance, the least able students may disproportionately take courses from the instructors with the best student evaluations, biasing downward our estimates of learning for the best instructors. Third, our results for future grades may reflect selection into future classes. Recall that we only observe future grades for students who take subsequent economics classes. Students who do well in one economics class may be more likely to take future economics classes. If more highly rated professors make economics more attractive particularly for students with low economics ability, the relationship between grades and the number of future classes taken will be weaker for students taking classes from the highly rated professors. In this case, our future grades measure will be biased downward for highly rated instructors relative to less highly rated professors, leading us to underestimate the effect of learning on evaluations. We will examine this possibility. A fourth explanation is that students from more highly rated professors may be

induced into taking more difficult future classes. We will also examine this possibility. A fifth interpretation is that the costs to students in courses where they learn much may offset the benefits they perceive. Lastly, students may be unable to gauge how much they have learned in their classes. The weak relationship between grades in future courses and the item that specifically captures learning, suggests that the last explanation may be the right one. We investigate these explanations below.

Precision of our Leaning Measure

Our estimates of learning may be noisy because of sampling error. To address this possibility, we estimate the share of the variance in our estimate of learning in each section that is common to all students in the section as opposed to sampling error. To do this, we split each class into two equally-sized halves and calculate the covariance between learning in each half. Formally, let $\theta_{sj} = \mu_s + \varepsilon_{sj}$ denote our estimate of learning for portion $j \in \{1,2\}$ of section *s*, which equals the learning in section *s*, μ_s, plus sampling error in portion *j* of the section, ε_{sj}.

We estimate, $Cov(\theta_{s1}, \theta_{s2})^{\frac{1}{2}} = Var(\mu_s)^{\frac{1}{2}}$. This measure gives the variation in learning across sections because it represents the variation in future grades for students who took a particular section in the absence of any sampling error. We also calculate the share of our future grades measure that represents learning as opposed to sampling error by calculating

$$\frac{Cov(\theta_{s1}, \theta_{s2})}{Var(\theta_s)} = \frac{Var(\mu_s)}{Var(\mu_s) + Var(\varepsilon_s)}.$$

Here ε_s denotes sampling error in the entire section.

Second in our regressions (**) of future grades on future section dummy variables and base-section dummy variables, we test for the statistical significance of the base-section dummy variables (vector θ) which are our measure of base-section learning. Third, we regress the base-section dummy variables from (**) on instructor dummy variables. The second-stage model is given by

$$\theta_s = \phi I_s + u_s,$$

where I_s denotes a vector of dummy variables for the instructors teaching the base section. It seems reasonable to assume that learning varies across sections and across instructors. Under this assumption, we expect section dummy variables and instructor dummy variables to be statistically significantly related to our learning measure (i.e., future grades).

Table 4 reports results for the three courses. As shown in the top panel, there is substantial variation in learning across sections – the standard deviations range between .2 and .3 grade points. Moreover, between 74% and 88% of the variance in future grades is due to section level learning, so our estimates of learning are quite precise. When we estimate (**), including controls for section characteristics, F-tests for the joint significance of the base-section dummy variables soundly reject the null hypothesis that base-section grades are not important determinants of future grades. For all three courses, the P-values are less than .0001.

As shown in the lower panel, more than half of the section-learning effects for principles of macroeconomics and intermediate microeconomics are due to instructor effects. Instructors account for 44% of the variation in the section-learning effects for principles of microeconomics We also reject the null hypothesis of no instructor effects, with a P-value less than .0001 for macro-principles and with P-values of .001 for micro-principles and .015 for intermediate microeconomics.

Based on these results, we conclude that although they contain a small amount of measurement error, grades in future courses are a valuable measure of learning in base-sections. The substantial variations in learning across sections and the strong effect of instructors on learning are also noteworthy and indicate the importance of evaluating instructors based on the learning that they produce.

Selection

This section considers whether selection biases our estimates. There are a number of selection arguments. The most simple is that there may be selection into base sections, so that

variations in learning and grades are due to differences in student ability.[16] We have addressed this argument by including SAT and ACT scores for the students for whom these scores are available in our regressions (*) and (**). Doing so reduced the sample size and had little impact on the estimates. For principles of microeconomics, we have also restricted the sample for which we estimate learning to students who took principles of microeconomics in the Fall of their first year. These students presumably have little information about instructors. (This strategy is similar to Hoffmann and Oreopoulos [2006]. Results were less precise but similar to those presented above.

Selection into Future Classes

There are other selection arguments. For instance, the effect of future grades on evaluations may be biased downward because students with low ability in economics take more additional economics courses after taking a course from a highly rated instructor than after taking a course from a less highly rated instructor.[17] To test this hypothesis, we estimate Tobit models of the number of future courses that student *i* in section *s* takes, *Future Class*$_{is}$. Our first model is,

$$FutureClass_{is}^{*} = e_s\beta + X_{is}\Gamma + W_s\Pi + \varepsilon_{is}$$

$$FutureClass_{is} = \begin{cases} FutureClass_{is}^{*} \ if \ \ FutureClass_{is}^{*} \geq 0 \cdot \\ 0 \ if \ \ FutureClass_{is}^{*} < 0 \end{cases}$$

Here e_s denotes the evaluation in section *s*; X_{is} denotes student characteristics; and W_s denotes characteristics of the instructor and section. This model can be used to determine whether students take more economics classes after taking a class from a highly-rated instructors than they do after taking a class from a less highly-rated instructor, in which case $\hat{\beta} > 0$. Our second

[16] Another selection possibility is that students who expect to receive bad grades drop classes and therefore do not complete evaluations (Becker and Powers [2001]). If, within a class, students expecting lower grades give lower evaluations, self-selection would raise both the observed average course grade and the observed evaluation. Unfortunately, our data do not permit us to identify students who dropped a course.

[17] Alternatively, students who are more interested in economics may rate their instructors better and continue with economics classes even if they are not as capable. Random variations across sections in student motivation might produce more low-quality students going on to take more economics classes when ratings are higher. These

model is,

$$FutureClass_{is}^{*} = g_{is}\beta + g_{is}e_s\pi + X_{is}\Gamma + \Phi_s + \varepsilon_{is}$$

$$FutureClass_{is} = \begin{cases} FutureClass_{is}^{*} \ if \ \ FutureClass_{is}^{*} \geq 0 \cdot \\ 0 \ if \ \ FutureClass_{is}^{*} < 0 \end{cases}$$

As above, e_s denotes the evaluation in section *s* and X_{is} denotes student characteristics; g_{is} gives the grade received by student *i* in section *s* and Φ_s denotes a set of section dummy variables, which are estimated explicitly and account for differences across base sections in the probability of taking future courses. With section fixed effects, the instructor and section characteristics (including the direct effect of student evaluations) are captured by the section fixed effects. The parameter β gives the difference between the number of subsequent economics courses taken by students with higher grades relative to those with worse grades. The parameter π, on the interaction between grades and evaluations, is of particular interest. If $\pi > 0$ ($\pi < 0$), then the relationship between students' grades and the number of future courses taken is stronger (weaker) in sections with higher evaluations.

While one might have expected that students from sections with higher evaluations would be particularly likely to take additional economics classes, the estimates reported in the odd numbered columns of table 5 show little relationship between student evaluations and the number of subsequent economics classes taken. The estimates for principles of microeconomics and intermediate microeconomics are both positive but statistically insignificant, while the estimate for principles of macroeconomics is negative and statistically significant. Thus, there is no evidence that students take more classes after having more highly rated instructors. This finding is somewhat disturbing from a revealed preference perspective if student evaluations are supposed to capture the quality of instruction. Students who take classes from foreign-born instructors and lecturers are less likely to take future classes.

The estimates in the even-numbered columns of the table show a strong positive

---

estimates also test for this hypothesis.

relationship between grades in the current course and the number of subsequent economics classes taken for students in principles of microeconomics and principles of macroeconomics. The relationship is negative but insignificant for students in intermediate microeconomics courses. On the other hand, the relationship between grades and the probability of taking future courses does not depend on evaluations in any of the three types of classes. Here too, there is little evidence that selection explains the weak relationship between grades in subsequent classes and student evaluations.

We also estimate our learning measure using a formal selection model. For these estimates, we look at students who took principles of microeconomics as their first principles course and their grades in principles of macroeconomics. Our instruments for whether students take subsequent classes, which were excluded from the future grade equation, are a set of interactions between the college that housed the student's major at the time of enrollment in principles of microeconomics and time (and its square). This is a good instrument, because it reflects exogenous changes in the requirements of majors and advising practices. We included college dummy variables in the equations for taking principles of macroeconomics and in the grade equation for principles of macroeconomics, so the selection model is estimated from variations over time in the share of principles of microeconomics students taking principles of macroeconomics within majors.

The results, which are quite similar to those in table 2, are reported in Appendix Table 3. Consistent with the previous results, there is a strong relationship between student evaluations and grades, which is unaffected by the inclusion of learning. While learning measures are positively related to evaluations without controls for grades, the relationship becomes small and is statistically insignificant when grades are included in the regression equation.

The Difficulty of Future Courses Taken

Another selection argument focuses on the particular classes that students take. Students who take a course from a more highly rated professor may take additional classes that are more difficult than do students whose prior course is from a less highly-rated instructor. While our

estimates of learning based on future grades include course fixed effects, if a disproportionately large number of students from a particular class take a course that yields lower average grades for the same amount of achievement, it will lead us to underestimate learning from those sections.

Our data provide a convenient test for this hypothesis insofar as intermediate microeconomics (the third class taken by most students) is offered in two versions – a standard course and a calculus-based course, taken by roughly 13% of the students in our sample. For students who took an intermediate microeconomics class, we estimate

$$Hard\ Intermediate_{is} = e_s\beta + X_{is}\Gamma + \varepsilon_{is}.$$

Here $Hard\ Intermediate_{is}$ is a dichotomous variable equal to 1 if the person took the mathematical intermediate microeconomics class and 0 if the person took the less mathematical intermediate microeconomics class; e_s denotes the evaluation in section *s*, and X_{is} denotes the student characteristics. The parameter β indicates whether students who took sections with higher ratings were more or less likely to take the more mathematical intermediate class.

The estimates are reported in Table 6. For principles of microeconomics we find no relationship between student evaluations and the probability of taking the more mathematical intermediate microeconomics course. For principles of macroeconomics the estimates indicate that students in more highly rated sections are less likely to take the more mathematical intermediate microeconomics course.

These estimates indicate that our finding of no relationship between evaluations and learning is not due to this potential source of bias. Overall, we conclude that there is little evidence that selection can account for the weak relationship between evaluations and learning reported above.

VIII. Learning and Grades

Our estimates show that there is substantial variation in learning across sections and that instructor effects account for much of this variation. This section considers how observable

instructor characteristics are related to our learning measure. We also study how current grades are related to learning and instructor characteristics. These are estimates of equations (****.1) and (****.2). Given that controlling for current grades eliminates the relationship between learning and evaluations, we anticipate that learning and current course grades are positively correlated.

The results are reported in table 7. The first three columns report results for principles of microeconomics. They show no systematic relationship between instructor characteristics and current course grades. Students who took principles of microeconomics from a foreign-born instructor or a lecturer are found to do better in subsequent classes. As expected, we find that grades in the current course are positively related to grades in future courses. This finding is consistent with instructors giving higher grades to students who know more at the end of the course.

Results for principles of macroeconomics, reported in columns (4) through (6), show that none of the observable instructor characteristics are related to subsequent grades, but that women tend to give lower grades while graduate teaching associates and instructors with Ph.D.s tend to give higher grades. As shown in column (5), grades in later courses are strongly associated with higher grades in the current course, but the previous results are robust to controlling for grades in later courses.

Results for intermediate microeconomics, reported in columns (7) through (9) show that none of the instructor characteristics is statistically significantly related to current grades. Students who took intermediate microeconomics from a foreign-born instructor tend to receive lower grades in later courses. Otherwise, none of the observable instructor characteristics is statistically significantly related to grades in later courses. As before, the current course grades are positively related to grades in future courses.

It is noteworthy that we generally find large instructor effects, but that the estimates in Table 7 show no consistent relationship between observable instructor characteristics and grades in future courses. This finding parallels the literature on teacher effects in primary and secondary

schools, where teacher effects are found to be large, but observable teacher characteristics have only weak effects (see, for example, Rivkin, Hanushek, and Kain [2005]). Thus here, as in that literature, the characteristics of instructors that matter the most are unobservable.

IX. Optimal Evaluation Criteria

As indicated it is possible to construct alternative evaluation criteria based on learning and the course experience, which can be estimated from equation (3) provided that consistent estimates of unbiased estimates of $\hat{\phi}_g$ and $\hat{\phi}_h$, in (3) and (***) are available.[18] This section considers a range of alternative criteria. To do this, table 8 reports the correlation between the raw student evaluations and these other criteria, with a lower correlation implying a greater discrepancy between the raw student evaluations and the alternative criteria.

The second row in each panel show the course experience, estimated from (3). These estimates are based on the regression reported in column 8 of table 2 and Appendix Table 1. The correlation between these variables and the overall rating ranges from .7 to .9, so simply adjusting for grades, learning, section, student, and instructor characteristics has a substantial impact on ratings. The third row in each panel shows the correlations with learning . Given the preceding results, it is not surprising that these correlations are consistently low and frequently negative, but the implications of this result are striking. It says that a social planner who cared only about learning would want to disregard student evaluations (or, ignoring behavioral responses, reward instructors slightly for low evaluations!).

The last row in each panel assumes that the social planner places equal weight on the student's experience and learning. Here the correlations with evaluations range between .4 and .6, indicating that roughly half of the variation in evaluations is noise to a planner with this objective. We believe that these results are conservative insofar as it is likely that society places substantially more weight on learning than the course experience.

---

[18] As indicated, the evidence above suggests that students have little information about the amount that they have learned so that $\phi_h \approx 0$, leaving $\hat{\phi}_g$ consistent.

X. Conclusions

We have shown that student evaluations differ from the ideal construct, because they do not reflect learning and are sensitive to grading leniency. We have no reason to believe that the focus on current grades and uncertainty about learning is specific to the current setting. Thus, we expect results to be qualitatively similar in higher education more generally. Even if student evaluations did not suffer from these two problems, it is unlikely that SEI scores would properly weight learning relative to the course experience – they would do so only if the social planner places the same weight on these items as students. Moreover, there is evidence that evaluations vary with class characteristics, including the type of section and composition of the class, and some evidence that students give lower evaluations to women and to foreign-born instructors.

Guarding against grading leniency is relatively simple (at least when evaluations do not reflect learning as is the case in our data). One simply needs to adjust evaluations for current and future grades using a simple regression procedure. Similarly, it is straightforward to estimate learning from administrative records on performance in subsequent classes. If one is willing to choose how much weight should be placed on learning relative to the course experience, it is possible to construct instructor-performance measures that reflect both variables. Evaluations could easily be adjusted for class characteristics and, depending on whether one believes that the lower scores received by women and foreign-born instructors reflect gender- or ethnic-discrimination as opposed to unobserved teaching attributes, evaluations could also be adjusted for instructor characteristics.

Beyond the issues discussed here, any evaluation method will be affected by sampling error. Extremely high or low evaluations may be cause of concern, but moderate variations in between evaluation scores are not. Similarly large changes for instructors merit attention, but it is unclear if moderate changes contain useful information. We have identified problems with student evaluations and proposed a number of modifications. If they are used at all to measure teaching quality, it may be valuable to supplement them with peer reviews of teaching.

References

Abrami, Philip C.; Sylvia d'Apollonia; and Steven Rosenfield "The Dimensionality of Student Ratings of Instruction: What We Know and What We Do Not." in *Effective Teaching in Higher Education Research and Practice*, R. P. Perry and J. C. Smart, Eds. New York: Agathon Press.

Baird, John S., 1987. Perceived Learning in Relation to Student Evaluation of University Instruction. Journal of Educational Psychology 79, 90-91.

Basow, Susan A., 1995. Student Evaluation of College Professors: When Gender Matters. Journal of Educational Psychology 87, 656-665.

Becker, William E., and Michael Watts, 1999. "How Departments of Economics Evaluate Teaching." American Economic Review 89, 344-349.

Becker, William E. and John Powers. 2001. "Student Performance, Attrition, and Class Size Given Missing Student Data." *Economics of Education Review* 20, 377-88.

Beddard, Kelly and Peter Kuhn. 2005. "Where Class Size Really Matters: Class Size and Student Ratings of Instructor Effectiveness." Working Paper.

Bettinger, Eric, Long, Bridget Terry, 2004. Do College Instructors Matter? NBER Working Paper Series No. 10370. Cambridge, Massachusetts: National Bureau of Economic Research.

Boex, L. F. Jameson, 2000. Attributes of Effective Economics Instructors: An Analysis of Student Evaluations. Journal of Economic Education 31, 211-227.

Bollinger, Christer R., Hoyt, Gail Mitchell, McGoldrick, KimMarie, 2005a. Attitude, Performance and Gender in Economics Principles Courses. Working paper. Department of Economics, University of Kentucky, Lexington, KY 40506.

_____, 2005b. Save those Clippings, but Leave the Computer Off: The Efficacy of Media Use in the Classroom. Department of Economics, University of Kentucky, Lexington, KY 40506.

Bosshardt, William, Watts, Michael, 2001. Comparing Student and Instructor Evaluations of Teaching. Journal of Economic Education 32, 3-17.

Cohen, Peter A. 1981. "Student Ratings of Instruction and Student Achievement: A Meta-Analysis of Multisection Validity Studies." *Review of Educational Research* 51 (No. 3, Fall): 281-309.

DeCanio, Stephen J., 1986. Student Evaluations of Teaching—A Multinomial Logit Approach. Journal of Economic Education 17, 165-176.

Dowell, David A. and James A Neal. 1982. "A Selective Review of Student Ratings of Teaching." *The Journal of Higher Education* 53 (No. 1, Jan.-Feb.): 51-62.

Ehrenberg, Ronald G., 2004. Prospects in the Academic Labor Market for Economists. Journal of Economic Perspectives 18, 227-238.

Feldman, Kenneth A. 1997. "Identifying Exemplary Teachers and Teaching: Evidence from Student Ratings" in *Effective Teaching in Higher Education Research and Practice*, R. P. Perry and J. C. Smart, Eds. New York: Agathon Press.

Golden, Daniel. 2006. "Colleges, Accreditors Seek Better Ways to Measure Learning." *The Wall Street Journal*. Monday, November 13, 2006. B1.

Grimes, Paul W., Millea, Meghan J., Woodruff, Thomas W., 2004. Grades—Who's to Blame? Student Evaluation of Teaching and Locus of Control. Journal of Economic Education 35, 129-147.

Greenwald, Anthony G. and Gerald M. Gillmore. 1997. "Grading Leniency is a Removable Contaminant of Student Ratings." *American Psychologist* 52 (No. 11): 1209-1217.

Hoffmann, Florian and Philip Oreopoulos. 2006. "Professor Qualities and Student Achievement." Working Paper.

Kelley, Allen C., 1972. Uses and Abuses of Course Evaluations as Measures of Educational Output. Journal of Economic Education 4, 13-18.

Krautmann, Anthony, Sander, William, 1997. Grades and Student Evaluations of Teachers. Economics of Education Review 18, 59-63.

Marsh, Herbert W. "Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Utility." *Journal of Educational Psychology* 76 (No. 5): 707-754.

Marsh, Herbert W. "Students' Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research." *International Journal of Educational Research* 11: 253-388.

Marsh, Herbert W. "Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness." Working Paper.

Marsh, Herbert W., Roche, Lawrence A., 2000. Effects of Grading Lenciency and Low

Workload on Students' Evaluations of Teaching: Popular Myth, Bias, Validity, or Innocent Bystanders? Journal of Educational Psychology 92, 202-228.

McCulloch, J. Huston, 1998. Deflating the SEI for Grade Inflation. Working Paper, Department of Economics, The Ohio State University, Columbus OH 43210.

McKenzie, Richard B., 1975. The Economic Effects of Grade Inflation on Instructor Evaluations: A Theoretical Approach. Journal of Economic Education 6, 99-105.

Mirus, Rolf, 1975. Some Implications of Student Evaluations of Teachers. Journal of Economic Education 5, 35-37.

Nelson, Jon P., Lynch, Kathleen A., 1984. Grade Inflation, Real Income, Simultaneity, and Teaching Evaluations. Journal of Economic Education 15, 21-37.

Nichols, Alan, Soper, John C., 1972. Economic Man in the Classroom. Journal of Political Economy 80, 1069-1073.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain, 2005. Teachers, Schools, and Academic Achievement. Econometrica 7, 417-458.

Sarwark, S., Smith, J., MacCallum, R., Cascllar, E. C., 1995. A Study of Characteristics of the SPEAK Test. RR 94047. Princeton, NJ, Educational Testing Service.

Sheets, D. F., Topping, E. E., 2000. Assessing the Quality of Instruction in University Economics Courses: Attrition as a Source of Self-Selection Bias in Mean Test Scores. The Journal of Economics 26, 11-21.

Sheets, Doris F., Topping, Elizabeth E., Hoftyzer, John, 1995. The Relationship of Student Evaluations of Faculty to Student Performance on a Common Final Examination in the Principles of Economics Course. The Journal of Economics 21, 55-64.

Siegfried, John J., Kennedy, Peter E., 1995. Does Pedagogy Vary with Class Size in Introductory Courses? American Economic Review 85, 347-351.

Shmanske, Stephen, 1988. On the Measurement of Teacher Effectiveness. Journal of Economic Education 19, 307-314.

Theall, Michael and Kenneth A. Feldman. 2006. "Commentary and Update on Feldman's (1997) 'Identifying Exemplary Teachers and Teaching: Evidence from Student Ratings.'" Working Paper.

United States Department of Education. 2006. *A Test of Leadership: Charting the Fugure of U.S. Higher Education.* Washington, D.C. 2006.

Watts, Michael, Bosshardt, William, 1991. How Instructors Make a Difference: Panel Data Estimates from Principles of Economics Courses. The Review of Economics and Statistics 85, 336-351.

Watts, Michael, Lynch, Gerald J., 1989. The Principles Courses Revisited. American Economic Review 79, 236-241.

White, Lawrence J., 1995. Efforts by Departments of Economics to Assess Teaching Effectiveness: Results of an Informal Survey. Journal of Economic Education 26, 81-85.

Figure 1. Instructors' choice of human capital and course experience when instructors get utility from challenging content.

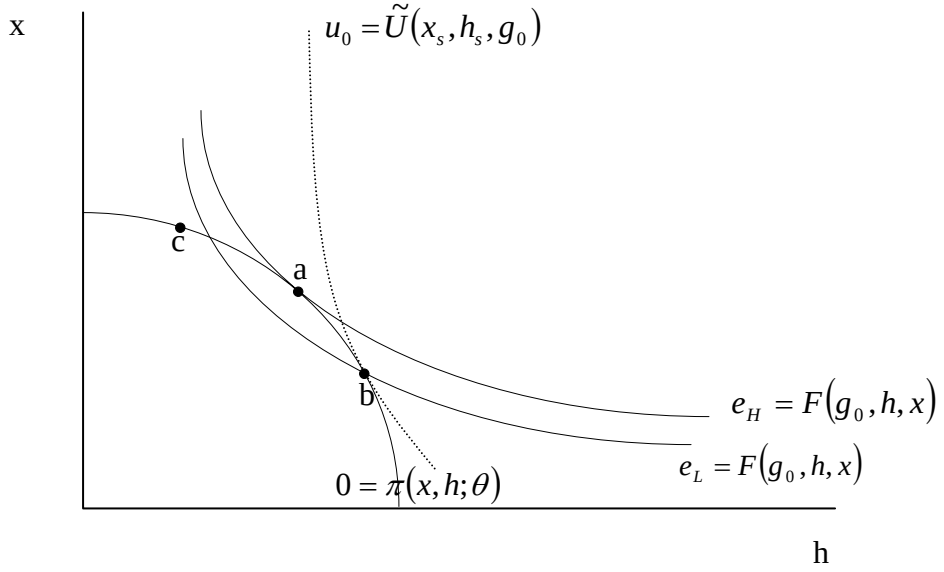

Table 1: Sample Characteristics

Variables	Prin. Micro.		Prin. Macro.		Inter. Micro.		Variable Definition
Number of Sections	194		122		88		
Number of Students	28,172		15,809		4,428		
Quality of Instruction	**Mean**	**S. D.**	**Mean**	**S. D.**	**Mean**	**S. D.**	
Overall	3.86	(0.44)	3.72	(0.54)	3.81	(0.64)	overall rating of instruction quality
Well Organized	4.13	(0.32)	4.01	(0.39)	4.06	(0.44)	instructor well organized
Intellectually Stimulating	3.49	(0.44)	3.51	(0.45)	3.69	(0.48)	intellectually stimulating
Interested in Teaching	4.16	(0.36)	4.01	(0.46)	4.15	(0.51)	instructor interested in teaching
Independent Thinking	3.76	(0.43)	3.66	(0.46)	3.88	(0.49)	encouraged independent thinking
Well Prepared	4.26	(0.35)	4.11	(0.41)	4.22	(0.43)	instructor well prepared
Helping	3.87	(0.46)	3.75	(0.53)	3.97	(0.59)	instructor interested in helping students
Learned Greatly	3.52	(0.49)	3.44	(0.55)	3.57	(0.64)	learned greatly from instructor
Learning Atmosphere	3.76	(0.43)	3.64	(0.52)	3.74	(0.62)	created learning atmosphere
Communicated Clearly	3.75	(0.49)	3.58	(0.59)	3.61	(0.73)	communicated subject matter clearly
Grades							
Current Course Grade	2.62	(0.30)	2.71	(0.28)	2.73	(0.34)	current grade, the mean grade the instructor gives in the current class (grade leniency)
Later Course Grade	2.81	(0.23)	2.85	(0.25)	2.83	(0.22)	future grade, the mean grades the students get in future Economics courses (human capital production)
Instructor Characteristics							
Instructor: Female	0.32	(0.47)	0.09	(0.28)	0.15	(0.36)	the instructor is female
Instructor: Foreign Born	0.16	(0.37)	0.21	(0.41)	0.33	(0.47)	the instructor is foreign born
Instructor: Lecturer	0.22	(0.41)	0.22	(0.42)	0.17	(0.38)	the instructor is a lecturer
Instructor: Grad. Associate	0.12	(0.32)	0.16	(0.36)	0.23	(0.43)	the instructor is a graduate teaching associate
Instructor: Has Ph.D.	0.80	(0.40)	0.75	(0.43)	0.77	(0.43)	the instructor has a PhD degree
Instructor: Years since Ph.D.	21.43	(13.33)	19.72	(16.82)	10.46	(11.45)	years since PhD
Instructor: Years at Institution	15.96	(11.98)	16.41	(14.35)	4.76	(5.17)	years hired by institution
Student Characteristics							
Students: Share Female	0.46	(0.06)	0.38	(0.06)	0.30	(0.06)	the portion of female students in the class
Students: Share Black	0.08	(0.03)	0.07	(0.03)	0.06	(0.04)	the portion of Black students in the class
Students: Share Hispanic	0.02	(0.01)	0.02	(0.01)	0.01	(0.02)	the portion of Hispanic students in the class
Class Characteristics							
Multi-Section Class	0.68	(0.47)	0.66	(0.48)	N/A	N/A	the class is a large lecture with multiple small recitation sections
Honors Class	0.06	(0.24)	0.07	(0.26)	N/A	N/A	the class is a honors class
Night Class	0.14	(0.35)	0.13	(0.34)	0.26	(0.44)	the class is a night class
Calculus Class	N/A	N/A	N/A	N/A	0.12	(0.33)	the course is calculus based
Response Rate	0.54	(0.21)	0.54	(0.19)	0.62	(0.14)	the response rate of SEI survey
Course Offered 1995	0.02	(0.13)	0.03	(0.18)	0.10	(0.30)	the course is offered in 1995
Course Offered 1996	0.05	(0.21)	0.09	(0.28)	0.10	(0.30)	the course is offered in 1996
Course Offered 1997	0.11	(0.32)	0.11	(0.31)	0.11	(0.32)	the course is offered in 1997
Course Offered 1998	0.08	(0.27)	0.12	(0.33)	0.10	(0.30)	the course is offered in 1998
Course Offered 1999	0.13	(0.34)	0.09	(0.29)	0.08	(0.28)	the course is offered in 1999
Course Offered 2000	0.14	(0.35)	0.15	(0.36)	0.07	(0.26)	the course is offered in 2000
Course Offered 2001	0.17	(0.38)	0.11	(0.31)	0.10	(0.31)	the course is offered in 2001
Course Offered 2002	0.18	(0.38)	0.14	(0.35)	0.14	(0.35)	the course is offered in 2002
Course Offered 2003	0.08	(0.28)	0.12	(0.33)	0.11	(0.31)	the course is offered in 2003
Course Offered 2004	0.04	(0.19)	0.04	(0.19)	0.08	(0.28)	the course is offered in 2004

Note: Standard deviations in parentheses.

Table 2: Determinants of SEI Overall Rating – Principle Microeconomics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Current Course Grade	0.29***		0.29***		0.30***	0.34***	0.34***	0.40***	0.40***	0.40***
	(0.06)		(0.07)		(0.07)	(0.07)	(0.10)	(0.11)	(0.13)	(0.13)
Learning		0.17**	-0.01		-0.02	0.03	0.01	-0.03	-0.05	-0.05
		(0.07)	(0.08)		(0.08)	(0.08)	(0.08)	(0.10)	(0.12)	(0.12)
Lag of Current Course Grade										0.02
										(0.08)
Instructor: Female				-0.28*	-0.25*			-0.28*	-0.07	-0.08
				(0.15)	(0.15)			(0.14)	(0.25)	(0.25)
Instructor: Foreign Born				-0.21	-0.22			-0.25	-0.44	-0.45
				(0.17)	(0.16)			(0.16)	(0.29)	(0.28)
Instructor: Lecturer				0.03	0.10			0.02	-0.08	-0.06
				(0.27)	(0.26)			(0.26)	(0.40)	(0.40)
Instructor: Grad. Associate				0.04	0.09			-0.04	0.12	0.14
				(0.54)	(0.51)			(0.53)	(0.77)	(0.76)
Instructor: Has Ph.D.				-0.05	-0.12			-0.20	-0.30	-0.29
				(0.56)	(0.53)			(0.53)	(0.84)	(0.82)
Instructor: Years since Ph.D.				0.00	0.01			0.01	0.01	0.01
				(0.01)	(0.01)			(0.01)	(0.02)	(0.02)
Instructor: Years at Institution				0.00	0.00			0.00	0.00	0.00
				(0.01)	(0.01)			(0.01)	(0.02)	(0.02)
Students: Share Female						-0.42		-0.28	-0.42	-0.42
						(0.28)		(0.29)	(0.34)	(0.34)
Students: Share Black						1.78**		1.95**	1.30	1.31
						(0.75)		(0.83)	(0.97)	(0.98)
Students: Share Hispanic						2.11		2.05	1.84	1.86
						(1.31)		(1.35)	(1.52)	(1.54)
Multi-Section Class							0.06	-0.07	-0.07	-0.07
							(0.12)	(0.15)	(0.17)	(0.17)
Honors Class							0.05	0.10	0.14	0.14
							(0.11)	(0.13)	(0.15)	(0.16)
Night Class							0.16**	0.16**	0.20**	0.20**
							(0.07)	(0.07)	(0.09)	(0.10)
Response Rate								-0.04	-0.09	-0.09
								(0.20)	(0.25)	(0.25)
Course Offered 1996								-0.01		
								(0.22)		
Course Offered 1997								-0.15	-0.01	-0.01
								(0.23)	(0.26)	(0.26)
Course Offered 1998								0.10	0.14	0.13
								(0.24)	(0.27)	(0.27)
Course Offered 1999								-0.01	0.11	0.11
								(0.23)	(0.26)	(0.26)
Course Offered 2000								0.00	0.08	0.07
								(0.23)	(0.26)	(0.26)
Course Offered 2001								-0.09	0.05	0.04
								(0.23)	(0.26)	(0.26)
Course Offered 2002								-0.03	0.14	0.12
								(0.24)	(0.27)	(0.27)
Course Offered 2003								0.17	0.32	0.31
								(0.25)	(0.28)	(0.28)
Course Offered 2004								-0.09	0.07	0.05
								(0.25)	(0.28)	(0.29)
Constant	4.06***	3.61***	4.08***	4.00***	4.25***	4.10***	4.04***	4.41***	4.42***	4.42***
	(0.08)	(0.11)	(0.16)	(0.56)	(0.54)	(0.21)	(0.17)	(0.64)	(0.98)	(0.97)
Number of Sections	194	194	194	194	194	194	194	194	135	135
R-Square	0.18	0.02	0.18	0.02	0.20	0.15	0.24	0.31	0.47	0.48

Note: Standard errors in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3: Determinants of Individual SEI Ratings – Principle Microeconomics

	Well Organized	Intellectually Stimulating	Interested in Teaching	Independent Thinking	Well Prepared	Helping	Learned Greatly	Learning Atmosphere	Communicated Clearly	Overall
Current Course Grade	0.14	0.33***	0.28***	0.31***	0.21**	0.35***	0.45***	0.36***	0.42***	0.40***
	(0.09)	(0.09)	(0.09)	(0.09)	(0.11)	(0.10)	(0.11)	(0.10)	(0.12)	(0.11)
Learning	0.01	0.05	0.01	-0.04	-0.05	-0.06	-0.02	-0.01	-0.03	-0.03
	(0.09)	(0.09)	(0.08)	(0.08)	(0.10)	(0.09)	(0.10)	(0.10)	(0.11)	(0.10)
Instructor: Female	-0.10	-0.20*	-0.14	-0.20	-0.15	-0.03	-0.28**	-0.23*	-0.24	-0.28*
	(0.11)	(0.11)	(0.11)	(0.14)	(0.12)	(0.13)	(0.14)	(0.13)	(0.16)	(0.14)
Instructor: Foreign Born	-0.23*	-0.26**	-0.08	-0.14	-0.21*	-0.14	-0.28*	-0.31**	-0.48***	-0.25
	(0.12)	(0.12)	(0.12)	(0.15)	(0.13)	(0.14)	(0.15)	(0.14)	(0.18)	(0.16)
Instructor: Lecturer	-0.06	-0.16	0.09	0.05	0.23	0.12	-0.05	0.07	0.02	0.02
	(0.21)	(0.20)	(0.21)	(0.26)	(0.22)	(0.24)	(0.25)	(0.24)	(0.31)	(0.26)
Instructor: Grad. Associate	0.20	-0.65	0.21	-0.25	0.57	0.61	-0.36	0.09	-0.03	-0.04
	(0.42)	(0.40)	(0.41)	(0.52)	(0.43)	(0.48)	(0.50)	(0.49)	(0.62)	(0.53)
Instructor: Has Ph.D.	0.07	-0.65	0.20	-0.21	0.38	0.37	-0.48	-0.12	-0.28	-0.20
	(0.42)	(0.40)	(0.41)	(0.52)	(0.43)	(0.48)	(0.50)	(0.49)	(0.62)	(0.53)
Instructor: Years since Ph.D.	0.00	0.01	0.00	0.01	-0.01	0.00	0.01	0.00	0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Instructor: Years at Institution	0.00	-0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Students: Share Female	-0.14	-0.37	-0.30	-0.19	-0.51*	-0.27	-0.52*	-0.27	-0.22	-0.28
	(0.26)	(0.26)	(0.25)	(0.24)	(0.30)	(0.28)	(0.30)	(0.28)	(0.32)	(0.29)
Students: Share Black	1.01	2.07***	1.72**	1.43**	0.96	1.78**	1.73**	1.90**	1.90**	1.95**
	(0.73)	(0.73)	(0.70)	(0.68)	(0.84)	(0.80)	(0.84)	(0.80)	(0.90)	(0.83)
Students: Share Hispanic	1.92	1.60	1.84	0.79	1.98	3.00**	2.56*	2.13	2.61*	2.05
	(1.18)	(1.19)	(1.14)	(1.10)	(1.36)	(1.30)	(1.37)	(1.30)	(1.46)	(1.35)
Multi-Section Class	0.11	-0.10	-0.06	-0.33***	0.06	-0.09	-0.06	-0.04	0.06	-0.07
	(0.13)	(0.13)	(0.12)	(0.12)	(0.14)	(0.14)	(0.15)	(0.14)	(0.16)	(0.15)
Honors Class	0.19*	0.22**	0.15	0.09	0.18	0.24**	0.13	0.10	0.04	0.10
	(0.11)	(0.11)	(0.11)	(0.11)	(0.13)	(0.12)	(0.13)	(0.12)	(0.14)	(0.13)
Night Class	0.11*	0.20***	0.15**	0.09	0.14*	0.18**	0.21***	0.16**	0.16**	0.16**
	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)
Response Rate	-0.08	0.01	-0.02	-0.26	-0.13	0.00	-0.03	-0.06	0.08	-0.04
	(0.17)	(0.17)	(0.17)	(0.17)	(0.19)	(0.19)	(0.20)	(0.19)	(0.22)	(0.20)
Course Offered 1996	0.01	0.26	0.06	0.06	0.04	0.12	0.03	0.02	-0.12	-0.01
	(0.19)	(0.20)	(0.19)	(0.18)	(0.22)	(0.21)	(0.22)	(0.21)	(0.24)	(0.22)
Course Offered 1997	0.02	0.04	-0.07	-0.13	0.04	-0.09	-0.15	-0.07	-0.28	-0.15
	(0.20)	(0.20)	(0.19)	(0.19)	(0.22)	(0.22)	(0.23)	(0.22)	(0.25)	(0.23)
Course Offered 1998	0.16	0.29	0.05	0.10	0.19	0.16	0.12	0.21	-0.05	0.10
	(0.20)	(0.21)	(0.20)	(0.20)	(0.23)	(0.23)	(0.24)	(0.23)	(0.26)	(0.24)
Course Offered 1999	0.07	0.22	0.01	0.07	0.06	0.13	0.01	0.08	-0.20	-0.01
	(0.20)	(0.20)	(0.19)	(0.19)	(0.22)	(0.22)	(0.23)	(0.22)	(0.25)	(0.23)
Course Offered 2000	0.05	0.21	-0.05	0.05	-0.01	0.11	0.02	0.08	-0.16	0.00
	(0.20)	(0.20)	(0.19)	(0.19)	(0.23)	(0.22)	(0.23)	(0.22)	(0.25)	(0.23)
Course Offered 2001	-0.01	0.19	-0.05	-0.01	-0.10	0.06	-0.04	0.06	-0.24	-0.09
	(0.20)	(0.20)	(0.20)	(0.20)	(0.23)	(0.22)	(0.24)	(0.23)	(0.26)	(0.23)
Course Offered 2002	0.08	0.21	-0.05	0.03	0.05	0.07	-0.01	0.11	-0.20	-0.03
	(0.20)	(0.21)	(0.20)	(0.20)	(0.23)	(0.23)	(0.24)	(0.23)	(0.26)	(0.24)
Course Offered 2003	0.20	0.34	0.12	0.21	0.14	0.28	0.23	0.27	0.05	0.17
	(0.21)	(0.21)	(0.20)	(0.21)	(0.24)	(0.23)	(0.25)	(0.23)	(0.27)	(0.25)
Course Offered 2004	0.07	0.24	0.00	0.00	-0.05	0.04	0.01	0.05	-0.17	-0.09
	(0.22)	(0.22)	(0.21)	(0.21)	(0.25)	(0.24)	(0.25)	(0.24)	(0.28)	(0.25)
Constant	4.10***	4.30***	4.22***	4.55***	4.17***	3.72***	4.58***	4.14***	4.30***	4.41***
	(0.52)	(0.51)	(0.51)	(0.60)	(0.56)	(0.58)	(0.62)	(0.60)	(0.73)	(0.64)
Number of Sections	194	194	194	194	194	194	194	194	194	194
R-Square	0.25	0.61	0.20	0.37	0.28	0.28	0.48	0.39	0.43	0.31

Note: Standard errors in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4: Tests of Grades in Future Sections as Measures of Base-Section Learning.

Base Course:	(1) Micro-Principles	(2) Macro-Principles	(3) Intermediate Micro
Standard Deviation in Learning across Sections			
Standard deviation	.249	.296	.215
Relative to variation in future grades across sections	.826	.876	.735
Test for Joint Significance of Base-Section Dummy Variables as Determinants of Grades in Future Courses, regression (**)			
F-Statistic	6.14	6.38	4.69
Degrees of Freedom (Numerator, Denominator)	(334, 34681)	(300, 21373)	(152, 10706)
P-Value	<.0001	<.0001	<.0001
Base-Section Instructors as Determinants of Grades in Future Courses			
F-Statistic for Joint Significance of Instructor-Effects	1.99	5.46	1.95
Degrees of Freedom (Numerator, Denominator)	(54, 139)	(33, 88)	(31, 56)
P-Value	.001	<.0001	.015
R^2 of Instructor Effects	.44	.67	.52

For principles of microeconomics, the F-statistic with 334 and 34681 degrees of freedom is 6.14, yielding a P-value beneath .0001. For principles of macroeconomics, the F-statistic with 300 and 21373 degrees of freedom is 6.38, yielding a P-value beneath .0001. For intermediate microeconomics, the F-statistic with 152 and 10706 degrees of freedom is 4.69, yielding a P-value beneath .0001.

Table 5: Determinants of the Number of Other Economics Courses Subsequently Taken

	Principles of Micro		Principles of Macro		Intermediate Micro	
SEI Overall Rating	0.05		-0.32***		0.18	
	(0.06)		(0.08)		(0.14)	
Course Grade		0.65***		0.89***		-0.45
		(0.20)		(0.21)		(0.46)
SEI Overall Rating * Course Grade		0.00		-0.02		0.09
		(0.05)		(0.06)		(0.12)
Student: Female	-0.93***	-0.85***	-0.63***	-0.57***	-0.95***	-0.96***
	(0.04)	(0.04)	(0.06)	(0.06)	(0.17)	(0.17)
Student: Black	-0.06	0.29***	-0.44***	-0.03	0.38	0.30
	(0.08)	(0.08)	(0.11)	(0.11)	(0.32)	(0.32)
Student: Hispanic	0.24*	0.44***	-0.13	0.16	0.37	0.15
	(0.15)	(0.14)	(0.21)	(0.21)	(0.65)	(0.64)
Multi-Section Class	0.52***		-1.09***			
	(0.11)		(0.14)			
Honors Class	1.29***		0.76***			
	(0.12)		(0.15)			
Night Class	0.16*		-0.21*		-1.26***	
	(0.09)		(0.12)		(0.21)	
Calculus Class					3.77***	
					(0.26)	
Instructor: Female	-0.18***		0.67***		0.12	
	(0.06)		(0.12)		(0.24)	
Instructor: Foreign Born	-0.25***		-0.87***		-0.31	
	(0.09)		(0.13)		(0.20)	
Instructor: Lecturer	-0.37***		-2.47***		0.67	
	(0.11)		(0.16)		(0.56)	
Instructor: Grad. Associate	-0.23		-2.78***		-0.04	
	(0.26)		(0.21)		(0.25)	
Instructor: Has Ph.D.	0.57**		-0.86***			
	(0.27)		(0.20)			
Instructor: Years since Ph.D.	-0.02***		(0.01)		0.04**	
	(0.01)		(0.01)		(0.02)	
Instructor: Years at Institution	0.01		-0.04***		-0.01	
	(0.01)		(0.01)		(0.03)	
Constant	-0.75**		4.06***		-0.99*	
	(0.38)		(0.46)		(0.56)	
Includes Section Fixed Effects		Yes		Yes		Yes
Number of Observations	26,666	26,666	14,729	14,729	4,111	4,111

Note: Standard errors in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 6: Determinants of Whether Subsequently Taking Calculus-Based Intermediate Microeconomics

	(1) Micro-Principles	(2) Macro-Principles
SEI Overall Rating	0.03	-0.04*
	(0.02)	(0.02)
Instructor: Female	0.02	-0.05***
	(0.02)	(0.02)
Instructor: Foreign Born	0.03	0.01
	(0.02)	(0.02)
Instructor: Lecturer	-0.03	0.00
	(0.04)	(0.03)
Instructor: Graduate Student	-0.05	0.04
	(0.07)	(0.04)
Instructor: Has Ph.D.	0.02	0.04
	(0.06)	(0.03)
Instructor: Years since Ph.D.	0.00	0.00
	(0.00)	(0.00)
Instructor: Years at Institution	0.00	0.00
	(0.00)	(0.00)
Student: Female	-0.02	-0.03***
	(0.02)	(0.01)
Student: Black	0.00	0.00
	(0.02)	(0.02)
Student: Hispanic	0.02	-0.01
	(0.03)	(0.04)
Multi-Section Class	-0.06	-0.10***
	(0.05)	(0.02)
Honors Class	0.20***	0.20***
	(0.05)	(0.03)
Night Class	0.00	-0.12***
	(0.04)	(0.02)
Course Offered 1996	-0.01	0.04**
	(0.04)	(0.02)
Course Offered 1997	0.03	0.05**
	(0.04)	(0.02)
Course Offered 1998	0.03	0.07***
	(0.04)	(0.02)
Course Offered 1999	0.05	0.09***
	(0.04)	(0.02)
Course Offered 2000	0.03	0.06**
	(0.03)	(0.02)
Course Offered 2001	0.05	0.07**
	(0.03)	(0.03)
Course Offered 2002	0.15***	0.18***
	(0.04)	(0.04)
Course Offered 2003	0.15***	0.19***
	(0.06)	(0.05)
Course Offered 2004	0.21	0.31***
	(0.22)	(0.08)
Constant	-0.02	0.23**
	(0.11)	(0.09)
Number of Sections	2188	2525
R-Square	0.07	0.09

Note: Robust standard errors in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 7: Determinants of Course Grades

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Principles of Microeconomics			Principles of Macroeconomics			Intermediate Microeconomics		
	Current Course Grade	Current Course Grade	Learning	Current Course Grade	Current Course Grade	Learning	Current Course Grade	Current Course Grade	Learning
Learning		0.17***			0.27***			0.23*	
		(0.06)			(0.07)			(0.13)	
Instructor: Female	-0.01	-0.01	0.04	-0.30**	-0.27**	-0.09	0.14	0.14	0.00
	(0.09)	(0.09)	(0.11)	(0.13)	(0.12)	(0.16)	(0.14)	(0.14)	(0.08)
Instructor: Foreign Born	0.10	0.06	0.24**	0.15	0.13	0.06	0.16	0.20	-0.11*
	(0.09)	(0.09)	(0.12)	(0.12)	(0.11)	(0.14)	(0.12)	(0.13)	(0.07)
Instructor: Lecturer	0.05	0.00	0.32*	0.26	0.19	0.26	-0.19	-0.09	-0.32
	(0.15)	(0.15)	(0.19)	(0.18)	(0.17)	(0.21)	(0.37)	(0.39)	(0.20)
Instructor: Grad. Associate	0.24	0.16	0.49	0.49*	0.41*	0.27	0.01	-0.02	0.11
	(0.30)	(0.30)	(0.38)	(0.26)	(0.25)	(0.31)	(0.16)	(0.16)	(0.09)
Instructor: Has Ph.D.	0.05	0.01	0.28	0.51*	0.54**	-0.13	-0.01	-0.01	0.00
	(0.31)	(0.31)	(0.40)	(0.29)	(0.27)	(0.34)	(0.01)	(0.01)	(0.01)
Instructor: Years since Ph.D.	0.00	0.00	0.01	-0.01	-0.01	0.01	0.01	0.02	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
Instructor: Years at Institution	0.00	0.00	0.01	0.02	0.02	0.00			
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)			
Multi-Section Class	0.03	0.07	-0.17	-0.09	-0.07	-0.07			
	(0.10)	(0.10)	(0.12)	(0.13)	(0.13)	(0.16)			
Honors Class	0.77***	0.69***	0.49***	0.61***	0.48***	0.48***			
	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.10)			
Night Class	-0.14**	-0.12**	-0.07	-0.06	-0.04	-0.06	0.13*	0.14*	-0.02
	(0.05)	(0.05)	(0.06)	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)	(0.06)
Calculus-Based Class							0.25***	0.19**	0.25***
							(0.08)	(0.08)	(0.06)
Constant	-1.18***	-1.28***	0.50	-0.55*	-0.49*	-0.18	-1.01***	-1.18***	0.65***
	(0.32)	(0.32)	(0.41)	(0.28)	(0.27)	(0.33)	(0.13)	(0.17)	(0.08)
Number of Sections	194	194	194	122	122	122	88	88	88
R-Square	0.65	0.66	0.35	0.66	0.70	0.57	0.32	0.31	0.34

Note: Standard errors in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 8: Correlations between Measures of Teacher Quality

	Student Evaluation	Course Experience	Learning (Regression Adjusted)	WFH Rating (50-50)
Principle Microeconomics				
Student Evaluation	1			
Course Experience	0.92	1		
Learning (Regression Adjusted)	-0.06	-0.10	1	
WFH Rating (50-50)	0.64	0.67	0.67	1
Principle Macroeconomics				
Student Evaluation	1			
Course Experience	0.69	1		
Learning (Regression Adjusted)	0.10	-0.11	1	
WFH Rating (50-50)	0.59	0.67	0.67	1
Intermediate Microeconomics				
Student Evaluation	1			
Course Experience	0.71	1		
Learning (Regression Adjusted)	-0.16	-0.07	1	
WFH Rating (50-50)	0.41	0.68	0.68	1

Note: The course experience is the residual of regression (8) in Table 2, which controls for human capital and grading leniency. Learning (regression adjusted) is the residual of a regression like that in column (3) of Table 3 but with only section characteristics and without instructor characteristics. WFH Rating = 0.5*Course Experience/SD(Course Experience) + 0.5*Learning/SD(Learning), where SD denotes the standard deviation of the relevant variable. Thus, it is a criterion where the course experience and learning receive equal weight.

Appendix Table 1: Determinants of SEI Overall Rating – Principle Macroeconomics and Intermediate Microeconomics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Principles of Macroeconomics										
Current Course Grade	0.36***		0.38***		0.39***	0.42***	0.47***	0.41**	0.92***	0.89***
	(0.11)		(0.14)		(0.14)	(0.14)	(0.16)	(0.19)	(0.24)	(0.23)
Learning		0.17	-0.03		-0.03	-0.01	0.00	-0.08	-0.14	-0.27
		(0.11)	(0.13)		(0.13)	(0.13)	(0.13)	(0.15)	(0.18)	(0.19)
Lag of Current Course Grade										0.38*
										(0.21)
Instructor: Female				-0.58**	-0.46			-0.35	0.23	0.39
				(0.29)	(0.28)			(0.25)	(0.31)	(0.31)
Instructor: Foreign Born				-0.21	-0.29			-0.43*	-0.30	-0.27
				(0.26)	(0.25)			(0.23)	(0.22)	(0.22)
Instructor: Lecturer				0.20	0.02			-0.26	-0.82**	-0.96***
				(0.40)	(0.39)			(0.35)	(0.32)	(0.32)
Instructor: Grad. Associate				-0.04	-0.27			-0.55	-1.26***	-1.49***
				(0.61)	(0.58)			(0.49)	(0.42)	(0.43)
Instructor: Has Ph.D.				-0.01	-0.29			-0.25	-1.70***	-2.26***
				(0.66)	(0.63)			(0.53)	(0.52)	(0.61)
Instructor: Years since Ph.D.				0.00	0.00			-0.01	0.05**	0.06***
				(0.02)	(0.02)			(0.02)	(0.02)	(0.02)
Instructor: Years at Institution				-0.01	-0.02			-0.03	-0.07***	-0.08***
				(0.02)	(0.02)			(0.02)	(0.02)	(0.02)
Number of Sections	122	122	122	122	122	122	122	122	86	86
R-Square	0.08	0.02	0.08	0.28	0.35	0.09	0.09	0.45	0.63	0.65
Intermediate Microeconomics										
Current Course Grade	0.74***		0.75***		0.82***	0.76***	0.53***	0.52**	1.06**	1.06**
	(0.19)		(0.20)		(0.22)	(0.20)	(0.19)	(0.22)	(0.43)	(0.45)
Learning		0.07	-0.15		-0.04	-0.14	-0.35	-0.41	-0.93**	-0.93*
		(0.26)	(0.25)		(0.26)	(0.25)	(0.24)	(0.26)	(0.47)	(0.48)
Lag of Current Course Grade										-0.01
										(0.44)
Instructor: Female				0.15	0.06			0.18	-0.86**	-0.86**
				(0.25)	(0.23)			(0.28)	(0.39)	(0.41)
Instructor: Foreign Born				-0.09	-0.22			-0.29	-0.97***	-0.97***
				(0.22)	(0.21)			(0.28)	(0.24)	(0.26)
Instructor: Lecturer				0.15	0.31			0.44	0.10	0.09
				(0.67)	(0.63)			(0.85)	(0.67)	(0.69)
Instructor: Has Ph.D.				-0.36	-0.41			-0.67*	-0.16	-0.16
				(0.29)	(0.26)			(0.36)	(0.38)	(0.39)
Instructor: Years since Ph.D.				0.00	0.01			0.00	0.01	0.01
				(0.03)	(0.02)			(0.03)	(0.02)	(0.02)
Instructor: Years at Institution				0.02	0.01			0.01	0.02	0.02
				(0.03)	(0.03)			(0.04)	(0.04)	(0.04)
Number of Sections	88	88	88	88	88	88	88	88	53	53
R-Square	0.10	0.00	0.12	0.04	0.24	0.13	0.18	0.37	0.65	0.65
Includes Student Characteristics						Yes		Yes	Yes	Yes
Includes Course Characteristics							Yes	Yes	Yes	Yes
Includes Response Rate								Yes	Yes	Yes
Includes Year Dummy Variables								Yes	Yes	Yes

Note: Standard errors in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table 2: Determinants of Individual SEI Ratings – Principle Macroeconomics and Intermediate Microeconomics

	Well Organized	Intellectually Stimulating	Interested in Teaching	Independent Thinking	Well Prepared	Helping	Learned Greatly	Learning Atmosphere	Communicated Clearly	Overall
Principles of Macroeconomics										
Current Course Grade	0.25*	0.53***	0.23	0.55***	0.23	0.43**	0.62***	0.50***	0.56**	0.41**
	(0.15)	(0.15)	(0.16)	(0.17)	(0.18)	(0.18)	(0.19)	(0.19)	(0.23)	(0.19)
Learning	0.01	-0.07	-0.03	-0.18	-0.12	-0.14	-0.09	-0.11	-0.13	-0.08
	(0.12)	(0.13)	(0.13)	(0.15)	(0.14)	(0.15)	(0.16)	(0.15)	(0.19)	(0.15)
Instructor: Female	-0.39*	-0.06	-0.20	0.06	-0.40*	-0.22	-0.22	-0.24	-0.39	-0.35
	(0.21)	(0.15)	(0.23)	(0.16)	(0.23)	(0.25)	(0.21)	(0.22)	(0.26)	(0.25)
Instructor: Foreign Born	-0.42**	-0.32**	-0.16	-0.23	-0.37*	-0.18	-0.47**	-0.44**	-0.84***	-0.43*
	(0.20)	(0.13)	(0.22)	(0.14)	(0.22)	(0.23)	(0.19)	(0.21)	(0.24)	(0.23)
Instructor: Lecturer	-0.27	-0.32*	-0.17	-0.49**	-0.24	-0.26	-0.29	-0.31	-0.20	-0.26
	(0.30)	(0.18)	(0.33)	(0.20)	(0.33)	(0.35)	(0.28)	(0.31)	(0.35)	(0.35)
Instructor: Grad. Associate	-0.27	-0.67***	-0.48	-0.95***	-0.35	-0.53	-0.58	-0.56	-0.41	-0.55
	(0.43)	(0.24)	(0.47)	(0.27)	(0.47)	(0.49)	(0.37)	(0.43)	(0.49)	(0.49)
Instructor: Has Ph.D.	-0.10	-0.61**	-0.19	-0.79**	-0.19	-0.31	-0.35	-0.33	-0.22	-0.25
	(0.47)	(0.28)	(0.51)	(0.31)	(0.51)	(0.54)	(0.42)	(0.47)	(0.54)	(0.53)
Instructor: Years since Ph.D.	-0.01	0.01	0.00	0.02**	-0.01	0.00	0.00	0.00	-0.01	-0.01
	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Instructor: Years at Institution	-0.02	-0.04***	-0.03*	-0.05***	-0.03*	-0.03*	-0.04**	-0.03**	-0.02	-0.03
	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Number of Sections	122	122	122	122	122	122	122	122	122	122
R-Square	0.34	0.64	0.46	0.56	0.42	0.46	0.57	0.50	0.47	0.45
Intermediate Microeconomics										
Current Course Grade	0.30*	0.25	0.28	0.53***	0.28	0.27	0.42**	0.51**	0.45*	0.52**
	(0.16)	(0.16)	(0.18)	(0.17)	(0.17)	(0.23)	(0.21)	(0.21)	(0.24)	(0.22)
Learning	-0.20	-0.08	-0.41*	-0.22	-0.33	-0.30	-0.27	-0.43*	-0.56**	-0.41
	(0.20)	(0.19)	(0.21)	(0.21)	(0.21)	(0.28)	(0.24)	(0.25)	(0.28)	(0.26)
Instructor: Female	0.19	0.18	0.25	0.08	0.13	0.34	0.31	0.31	0.34	0.18
	(0.19)	(0.19)	(0.23)	(0.21)	(0.20)	(0.26)	(0.28)	(0.27)	(0.31)	(0.28)
Instructor: Foreign Born	-0.24	-0.12	-0.16	-0.10	-0.11	-0.24	-0.29	-0.27	-0.60**	-0.29
	(0.19)	(0.19)	(0.23)	(0.20)	(0.20)	(0.26)	(0.28)	(0.27)	(0.30)	(0.28)
Instructor: Lecturer	0.24	0.08	0.10	0.65	0.21	-0.07	0.14	0.33	-0.14	0.44
	(0.57)	(0.58)	(0.69)	(0.62)	(0.60)	(0.78)	(0.85)	(0.82)	(0.93)	(0.85)
Instructor: Has Ph.D.	-0.22	-0.48*	-0.51*	-0.60**	-0.12	-0.63*	-0.71*	-0.71**	-0.64	-0.67*
	(0.25)	(0.25)	(0.30)	(0.27)	(0.26)	(0.34)	(0.37)	(0.35)	(0.40)	(0.36)
Instructor: Years since Ph.D.	-0.01	0.01	0.01	0.00	0.00	0.01	0.01	0.01	0.02	0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Instructor: Years at Institution	0.02	0.00	0.01	0.01	0.02	-0.01	0.00	0.01	-0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Number of Sections	88	88	88	88	88	88	88	88	88	88
R-Square	0.32	0.45	0.29	0.40	0.24	0.34	0.35	0.38	0.38	0.37

Note: Standard errors in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%. All regressions include student characteristics, class characteristics, the response rate, and year dummy variables.

Appendix Table 3: Determinants of SEI Overall Rating – Principle Microeconomics – Selection Corrected Learning Measure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Current Course Grade	0.28***		0.28***		0.28***	0.34***	0.28***	0.34***	0.44***	0.54***
	(0.06)		(0.07)		(0.07)	(0.07)	(0.10)	(0.11)	(0.15)	(0.17)
Learning (Selection Corrected)		0.13*	0.001		0.01	0.03	0.02	0.02	0.01	0.08
		(0.07)	(0.08)		(0.08	(0.08)	(0.08)	(0.09)	(0.13)	(0.16)
Lag of Current Course Grade										0.15
										(0.10)
Instructor: Female				-0.33**	-0.30*			-0.34**	-0.23	-0.28**
				(0.16)	(0.16)			(0.15)	(0.16)	(0.14)
Instructor: Foreign Born				-0.22	-0.24			-0.27	-0.70***	-0.84***
				(0.17)	(0.17)			(0.17)	(0.17)	(0.15)
Instructor: Lecturer				-0.01	0.08			0.01	-0.12	0.09
				(0.27)	(0.26)			(0.27)	(0.24)	(0.24)
Instructor: Grad. Associate				-0.1	0.03			-0.06	-0.23	-0.14
				(0.54)	(0.53)			(0.55)	(0.48)	(0.48)
Instructor: Has Ph.D.				-0.06	-0.07			-0.11	-0.31	-0.36
				(0.56)	(0.55)			(0.55)	(0.51)	(0.46)
Instructor: Years since Ph.D.				0.001	0.001			0.0005	-0.01	-0.01
				(0.01)	(0.01)			(0.01)	(0.01)	(0.01)
Instructor: Years at Institution				0.002	0.001			0.002	0.002	0.002
				(0.01)	(0.01)			(0.01)	(0.01)	(0.01)
Students: Share Female						-0.45		-0.31	-0.18	0.13
						(0.29)		(0.31)	(0.41)	(0.49)
Students: Share Black						1.60**		1.74*	1.00	1.48
						(0.8)		(0.89)	(1.16)	(1.39)
Students: Share Hispanic						1.76		2.04	1.34	1.33
						(1.35)		(1.40)	(1.79)	(2.13)
Multi-Section Class							0.08	-0.05	0.06	0.29
							(0.12)	(0.15)	(0.18)	(0.19)
Honors Class							0.13	0.14	0.16	0.35*
							(0.11)	(0.13)	(0.18)	(0.21)
Night Class							0.17**	0.17**	0.24**	0.37***
							(0.07)	(0.08)	(0.11)	(0.13)
Response Rate								-0.04	-0.24	-0.39
								(0.21)	(0.27)	(0.30)
Course Offered 1996								-0.19		
								(0.32)		
Course Offered 1997								-0.26	-0.01	-0.24
								(0.31)	(0.36)	(0.44)
Course Offered 1998								0.01	0.3	0.18
								(0.31)	(0.37)	(0.45)
Course Offered 1999								-0.11	0.32	0.24
								(0.30)	(0.35)	(0.43)
Course Offered 2000								-0.10	0.2	0.06
								(0.31)	(0.35)	(0.43)
Course Offered 2001								-0.19	0.19	0.04
								(0.31)	(0.35)	(0.43)
Course Offered 2002								-0.11	0.26	0.06
								(0.31)	(0.35)	(0.44)
Course Offered 2003								0.08	0.42	0.20
								(0.32)	(0.36)	(0.45)
Course Offered 2004								-0.14	0.23	-0.06
								(0.33)	(0.38)	(0.46)
Constant	3.88***	3.90***	3.88***	4.12***	4.10***	3.97***	3.81***	4.25***	4.45***	4.58***
	(0.06)	(0.08)	(0.08)	(0.56)	(0.55)	(0.17)	(0.09)	(0.68)	(0.72)	(0.77)
Number of Sections	183	183	183	183	183	183	183	183	128	128
R-Square	51	51	51	51	51	51	51	51	26	26

Note: Standard errors in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%. Our instruments for whether students take principles of macroeconomics are interactions between the college that housed the student's major at the time of enrollment in principles of

microeconomics and time (and its square). The reported estimates include dummy variables for the colleges.