Evaluating Teaching in Higher Education

September 2008

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

Belton M. Fleisher
The Ohio State University[*] and IZA
fleisher.1@osu.edu

Masanori Hashimoto
The Ohio State University[*]
Hashimoto.1@osu.edu

Abstract

This paper develops an original measure of learning in higher education, based on grades in subsequent courses. Using this measure of learning, this paper shows that student evaluations are positively related to current grades but unrelated to learning once current grades are controlled. It offers evidence that the weak relationship between learning and student evaluations arises, in part, because students are unaware of how much they have learned in a course. The paper concludes with a discussion of easily-implemented, optimal methods for evaluating teaching.

---

[*] Department of Economics; Ohio State University; 1945 North High Street; Columbus, OH 43210.

I. Introduction and Background

Given the difficulties associated with accurate assessments of teaching effectiveness, student evaluations traditionally have been the primary, if not the only, means of assessing teaching in higher education.[1] 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.

This paper develops a procedure for optimally assessing instructor quality that is easily implemented. This procedure applies basic regression techniques to data that are already available in machine-readable form at most or all institutions. It is based on a unique measure of the learning acquired in a section of a course based on the grades that the students in that section receive in subsequent courses in the same subject (adjusting for student characteristics). The availability of a learning measure allows us to investigate how students reward instructors for learning as well as for the grades the receive in their current course.

Economists and educational psychologists both have studied how grades and learning are related to student evaluations of teaching but in a way that prevents one from inferring whether students reward higher grades, more learning, or both. Instructors are expected to give higher grades to better performing students, making it necessary to control for both grades and learning when assessing the determinants of student evaluations. Surprisingly, we are unaware of studies that have considered both factors together. While the existing literature has not convincingly demonstrated that students reward learning and not grading leniency, even without these issues, student evaluations generate an incentive for instructors to inflate grades and almost surely place too little weight on learning.

There is a substantial literature on evaluations of teaching in educational psychology,

---

[1] White (1995) reports that U.S. economics departments predominantly use SET's to measure teaching effectiveness. He notes that there appears to be strong reluctance to rely on direct observation of teaching, particularly among research-oriented departments.

comprising thousands of items (Feldman [1997]). 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 raises questions about the extent to which these results will generalize to other courses. More importantly, this design is not suitable for determining how grades and learning separately affect evaluations because there is little if any variation in grades conditional on learning. Ironically, discussions with our students suggest that they often estimate how much they have learned in a course from the grade that they expect to receive.[2] 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 may be biased by grades. Because the multi-section design makes it virtually impossible to estimate the effect of grades and learning separately, educational psychologists have generally relied on indirect methods to address the effect of grading leniency (see, for example, Greenwald and Gillmore [1997]).[3]

The literature on the economics of education is less sanguine about the relationship between learning and evaluations. It has focused on how grades affect evaluations, providing evidence that high grades raise student evaluations and argued that the use of evaluations may lead to grade inflation.[4] Results from empirical work linking expected grades to evaluations are

---

[2] This view was expressed at a presentation of this work at our Undergraduate Economics Society at our university at which approximately 30 students were in attendance.

[3] 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.

[4] Becker and Watts (1999) criticize economics departments for "following the herd" in their uncritical use of SET 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 SET's on student learning and grade inflation. McKenzie (1975) develops a simple model of consumer

mixed.[5] This paper departs from this work in two ways. First, it uses actual course grades rather than expected grades. While students generally do not receive grades until after completing their evaluations, we prefer to use actual grades because expected grades are likely to be quite noisy and 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 possible signals from the instructor. Second, unlike most of the literature, this paper measures grades using the average grade in a section rather than individual-level grades. In an individual-level regression, most of the variation in grades arises from individual differences in grades within a section. Therefore, 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. Third, as indicated above, this paper is also the first to study how grades and learning are jointly related to evaluations.

We use our analysis of the determinants of student evaluations to suggest improved methods for evaluating instructors. Educational psychologists have argued that if there is little effect of grades on evaluations but there is a strong relationship between grades and learning, one would not want to adjust student evaluations for grades (Greenwood and Gillmore [1997]). If evaluations were affected by grades, however, it would be advisable to adjust evaluations for grades, provided that one can condition on learning. Such an adjustment would reduce instructor incentives to inflate grades and remove a source of bias in evaluations.

There are two additional factors that arise when relying on student evaluations. First student evaluations depend on students' assessment and valuation of how much they have

---

choice in which the use of SET'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).
[5] 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

learned in a course. Below, we provide evidence that student evaluators may not be well positioned to make that determination accurately. Even if students accurately assess their learning, they may place less weight than universities do 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 will distort instructors' incentives away from the social optimum.

This paper 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 also relates to a large literature in the economics of education on the determinants of student outcomes in primary and secondary education.

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.[6] In addition to information on student evaluations, the data contain all grades that students received in subsequent economics courses and rich information on student background, including race, gender, ethnicity, high school class rank and SAT and/or ACT scores. These data can be used to regression-adjust grades and our learning measure. The data show a strong positive relationship between student evaluations and both current grades and learning when these variables are included separately, but when these variables are included in the same model, the current grade is related to student evaluations but learning is not. There are many potential explanations for these results, including a variety of selection arguments. We devote considerable effort to five of them, concluding that, on average, students are not aware of how much they have learned in a class. There are no reasons to believe that the focus on current grades and uncertainty about learning is specific to economics or the

---

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

[6] 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.

institution studied and therefore our results are expected to generalize, at least qualitatively.

The availability of a measure of learning and the checks that we run on it entail a number of analyses that are of interest in their own right. For example, we also consider how instructor characteristics are related to measures of the quality of teaching. In some cases, female and foreign-born instructors receive lower student evaluations than male and US-born instructors. Learning, however, is not systematically related to instructor gender or national origin, nor are there 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 little is known about the effect of part-time and non tenure-track faculty on student learning and other measures of academic production. While we do not find that observable instructor characteristics are related to learning, we do find large variations across instructors in learning performance. This result is consistent with evidence from primary and secondary education (See Rivkin, Hanushek, and Kain [2005]).

We also investigate whether students in sections that rate their instructors more highly are more likely to take additional classes, a revealed-preference measure of quality (see Hoffmann and Oreopoulos [2006]). We find that student evaluations of teaching are unrelated to the number of subsequent economics classes that students take, further eroding our confidence in student evaluations.[7]

II. Data

The data set includes students who took principles of microeconomics, principles of macroeconomics or intermediate microeconomics at The Ohio State University between 1995 and 2004, and contains identifiers for the sections the students took, student demographic characteristics, and grades in all economics courses taken during this period. They also contain rich background information on students, including race, gender, ethnicity, high school class

---

[7] 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.

rank and SAT and/or ACT scores that can be used to regression-adjust grades and our learning measure. In the regressions, ACT scores are included for students for whom we have them and SAT scores are included for students for whom we do not have ACT scores (because ACT scores are available for more students). The estimates include a dummy variable for which score is included. We obtained data on all subsequent economics courses taken by these students through the end of academic year 2004. 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 is appropriate, as discussed above.

The evaluation instruments contain ten items, including an overall score, which is the principal measure of student evaluation used in this study. Other questions 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 shows the variable definitions and their means and standard deviations for the three sets of courses. The data set comprises 190 sections (with 26,666 students) in principles of microeconomics; 119 sections (with 14,729 students) in principles of macroeconomics; and 86 sections (with 4,111 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). The average course grade is close to 2.7 on a four-point scale (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.

III. Estimation

We employ a three-step strategy to estimate grades and learning and their relationship to student evaluations: (i) we first estimate grades (ii) then we estimate the amount of learning in each section based on grades in subsequent sections; (iii) finally we use these learning estimates to assess how grades and learning are related to evaluations. In addition, we analyze these

learning measures since they are of interest in their own right.

This section describes the procedure step-by-step in terms of principles of microeconomics, including how individual-level data on current and subsequent grades are collapsed to the section-level to be merged into section-level evaluations. Similar procedures were used for principles of macroeconomics and intermediate macroeconomics.

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, g_{is} is regressed on a vector of base section dummy variables $\vec{D}_{is}$ and the student's characteristics at the time of the base section, $\vec{X}_{is}$. The specification is

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

The coefficient ψ_s on the dummy variable for base section s gives the mean grade in the section adjusting for individual characteristics. These coefficients are used in the third stage to represent grades.

Step 2. Estimating Learning

Grades in subsequent courses are used to measure learning. 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. Grades in subsequent courses, g_{isj}, are regressed 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 student characteristics, $\vec{X}_{isj}$, at the time of section j. Formally,

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

The coefficient θ_s on the dummy variable for students who took base section s indicates how well these students do in later courses adjusting for their characteristics. This coefficient is our

7

measure of learning, or human capital produced in section s. The θ_s we estimate in this stage are used in our third step to control for learning produced in the section and the estimates are of interest in their own right.

Step 3. Evaluating Student Evaluations

Having estimated ψ_s, the grades in base section s and θ_s, the learning in base section s, we now regress the student evaluations for base section s, e_s, on learning, grades, and instructor and section characteristics, $\vec{W}_s$:

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

The coefficient ϕ_l indicates 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 and section characteristics are related to evaluations.

Additional Analyses

Once we obtain a measure of learning, we can estimate a variety of related effects. We investigate 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 by estimating,

$$\theta_s = \vec{W}_s' \vec{\beta}_3 + u_s. \qquad (4.1)$$

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,

$$\psi_s = \vec{W}_s' \vec{\beta}_4 + \theta_s \gamma + \xi_s. \qquad (4.2)$$

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

IV. Findings

Validity of the Learning Measure

We begin by validating the learning measure. The most obvious concern is that it is noisy because of sampling error. To address this possibility, we estimate the share of the variance in the estimate of learning in each section that is due to learning 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. Intuitively, the idiosyncratic components of learning in each half of the course will be unrelated to each other, and the covariance between them will give the variance in learning in the section as a whole. We now discuss the formal procedure.

Let $\theta_{sj} = \mu_s + \varepsilon_{sj}$ denote the 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 the covariance between the two randomly assigned halves of each section,

$$Cov(\theta_{s1},\theta_{s2})=Cov(\mu_{s1},\mu_{s2})+Cov(\mu_{s1},\varepsilon_{s2})+Cov(\varepsilon_{s1},\mu_{s2})+Cov(\varepsilon_{s1},\varepsilon_{s2}).$$

The terms involving the sampling errors, ε_{sj} drop out because they are orthogonal to the other components by construction, and the covariance between the μ_s gives its variance. Formally,

$Cov(\theta_{s1},\theta_{s2})^{\frac{1}{2}} = Var(\mu_s)^{\frac{1}{2}}$. This measure gives the standard deviation in learning across sections because it represents the variation in learning for students who took a particular section in the absence of any sampling error. We also calculate the share of our learning 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 regression (2), where we generate our learning measure by relating future grades to future section dummy variables and base-section dummy variables, we test for the statistical significance of the base-section dummy variables (vector θ) which measure base-section learning. Third, we regress the base-section dummy variables from (2) on instructor

9

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, section dummy variables and instructor dummy variables are expected to be statistically significantly related to the learning measure (i.e., future grades).

Table 2 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 .15 and .2 grade points. These differences imply that moving a student from a section with the mean level of learning to one with learning one standard deviation above the mean would raise his grades in all future economics classes from, say from a B- to more than half way to a B. Moreover, between 46% and 83% of the variance across base sections in the learning measure is due to learning at the section as opposed to noise, so our estimates of learning are quite precise. In other words, while (unobserved) individual factors such as motivation and ability account for the majority of the variation in future grades across *students*, only a minority (54% to 17%) of the variation in future grades across base *sections* is due to noise from sampling and the vast majority (46% and 83%) is signal.

When we estimate (5), 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 learning. 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 are due to instructor effects. Instructors account for 44% of the variation in the section-learning effects for intermediate microeconomics and 39% for principles of microeconomics. The null hypothesis of no instructor effects is also rejected with a P-value less than .0001 for macro-principles and with P-values of .01 for micro-principles and .1 for intermediate microeconomics. While much of the variation across sections is due to fixed

instructor effects, there is still considerable variation within instructors (due to variations in the rate of learning by doing or deterioration and idiosyncratic factors). For this reason, we focus on section-level (rather than instructor-level) estimates of learning for most of the analyses.

These results indicate that although they contain some sampling 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.

Principles of Microeconomics

This section reports estimation results for equation (3). We begin with 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. These estimates and others like them include instructor random effects.

The first column of table 3 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 the learning measure, which is also found to be positively associated with evaluations though with a smaller coefficient than current grades. When both current grades and learning are included in the same regression (column 3), the effect of the current grade dominates, and the coefficient on learning is small and insignificant.[8] We remind the reader that both current grades and the learning measure are regression adjusted for observable student characteristics.

The remaining columns examine a variety of other potential determinants of student evaluations. First, we include a set of instructor characteristics without controlling for grades or learning (column 4). Female instructors receive lower evaluations than men, as do foreign-born

---

[8] If learning is multi-dimensional, grades in the current course may capture learning of material that is valuable outside of future courses, including skills as an economic actor or a citizen. By measuring learning using grades in future classes, the present analysis will not capture these other dimensions of learning. An alternative explanation for the strong relationship between evaluations and current grades is that teachers who expect to receive bad

instructors, although these differences are not statistically significant in all specifications. 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, column (5) includes both current grades and learning along with instructor characteristics. Inclusion of these variables increases the gender and foreign-domestic gaps in evaluations slightly. 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 regression in column (6) includes 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 on the current course grade is significant and similar in magnitude. To summarize other statistically significant findings, column (6) shows that students in honors and evening sections give higher evaluations than other students (i.e. those in non-honors, daytime sections). Column (7) shows that the coefficients for honors 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 3 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 not statistically significantly related to evaluations in any of the regressions.

Our use of the actual current course grade as a measure of the expected grade in the

evaluations penalize their classes with harder exams or a harder curve.

12

course deserves some discussion. As noted already, 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 the regression the lagged grade – the mean grade in the last offering of the course by the instructor – along with the current grade. Column (8) presents results without the lagged grade for the sample for which the lagged grade is available. Including the lagged grade, in column (9) does not change the estimated coefficient of the current course grade or learning, 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

We use ten evaluative items, nine focusing on specific aspects of the course experience as well as the overall score, which has been the focus of the analysis thus far. Estimates for these individual items (not reported here, but available upon request) 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 eight of the ten cases. None of the evaluation items are statistically significantly related to learning.

One item directly measures learning, asking students whether they, "Learned greatly from instructor?" It is noteworthy that this measure is no more closely related to our learning measure based on 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 they expect to receive.

Principles of Macroeconomics and Intermediate Microeconomics

This section reports results for principles of macroeconomics and intermediate microeconomics as a check. There are only 60 percent as many macro-principles sections as there are micro-principles sections and fewer of the students in macro-principles take subsequent

classes. Similarly, there are fewer intermediate economics sections than micro-principles (less than half as many) or macro-principles (three quarters as many) and fewer students take later classes making the estimates noisier.[9] Thus, both sets of estimates will be more noisy than those for principles of microeconomics.

The results, which are reported in Appendix Table 1 in the same order as those for principles of microeconomics, are generally consistent with those for principles of microeconomics. For both principles of macroeconomics and intermediate microeconomics, grades in the current course are strongly related to student evaluations for later courses. In fact, the estimates are, if anything, stronger, than those for micro-principles. Learning is unrelated to evaluations once current grades are controlled. Again women and foreign born instructors tend to receive lower evaluations than men and domestic instructors, but these differences are not consistently statistically significant. Results, not reported here but available upon request, for individual evaluations items also show a strong relationship with current course grades but not learning.

Robustness

This section explores two robustness checks. One possibility is that learning in principles of microeconomics may be better captured by performance in other microeconomics classes. To explore this possibility we have estimated learning using grades in intermediate microeconomics only. The estimates are reported in the top panel of Table 4 and are similar to the previous estimates, with a strong positive relationship between current grades and student evaluations and a weak relationship between learning and student evaluations.

Because we are often interested in the effectiveness of particular instructors, we have also re-estimated our models using instructor-level data. To do this, we estimate the weighted mean of student evaluations and of our measures of grades and learning for each section taught by an

---

[9] 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 to estimate learning for micro-principles, while for macro-principles, learning can only be estimated for students who take a third economics course. Similarly, many business majors require intermediate microeconomics, but no additional classes.

instructor estimated in equations (1) and (2). In these calculations, each section was weighted by the number of students in that section. Equation (3) is then estimated at the instructor level with these instructor-level means as the independent and dependent variables[10]. (In the specifications that include section and instructor characteristics, the means of those variables are included.) These estimates, reported in the bottom panel of the table, show somewhat higher estimates of current grades than the section-level estimates, but the coefficient learning is small and statistically insignificant.

Summary

The highlights of what we found so far are:

1. There is a consistent positive relationship between grades in the current course and evaluations. This finding is robust to the inclusion of a wide range of controls and specifications.

2. There is no evidence of association 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.

V. Alternative Explanations

Through the various checks reported in Table 2, we feel confident that our measure of learning is valid. Our findings show that student evaluations are strongly related to grades and that learning, as measured by future grades, is unrelated to student evaluations once grades have been controlled. The obvious explanation for these findings is that grading leniency is an important determinant of evaluations and that students do not reward instructors who generate learning per se. There are, however, five other explanations that come to mind and this section

---

[10] We do not include the lagged evaluation because it is ill-defined when these regressions are run for instructors as

investigates them and shows that they are unlikely to explain our results.

First, the findings may indicate selection into courses – for example, 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. Second, the results for learning may reflect selection into future classes. Our learning measure can only be constructed for students who take subsequent economics classes. Better performing students in one economics course 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 whether students take additional classes will be weaker for students taking classes from the highly rated professors. In this case, our learning measure will be biased downward for highly rated instructors relative to less highly rated ones, leading us to underestimate the effect of learning on evaluations. We will examine this possibility. A third 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 fourth 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 our learning measure and the evaluation item that specifically addresses learning, suggests that the last explanation may be the most relevant one. We investigate these explanations below.

Selection into Base Sections

This section considers whether selection biases our estimates. There are a number of selection arguments. The simplest is that there may be selection into base sections, so that variations in learning and grades are due to differences in student ability.[11] As noted earlier, all the preceding estimates control for a rich set of student characteristics including ACT or SAT

---

opposed to individual courses. Also, lagged grades did not matter in the previous models.

[11] 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.

scores and high school class rank, thereby mitigating this concern. 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 example, the effect of learning, as measured by future grades, on evaluations may be biased downward if 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.[12] To test this hypothesis, we estimate logit models of whether student *i* takes subsequent economics courses, *Future Class$_{is}$*. The first model is,

$$FutureClass_{is} = \begin{cases} 1 \text{ if } e_s\beta + X_{is}\Gamma + W_s\Pi + \varepsilon_{is} > 0 \\ 0 \text{ if } e_s\beta + X_{is}\Gamma + W_s\Pi + \varepsilon_{is} \le 0 \end{cases}.$$

Here e_s denotes the evaluation in section *s*; X_{is} denotes student characteristics; and W_s denotes characteristics of the instructor and section. In addition to individual characteristics included above, in these regressions, X_{is} includes dummy variables for the college that houses the student's major and interactions between these dummy variables and a quadratic in time to account for (time-varying) differences in requirements to take economics classes across different units.[13] 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 $\hat{\beta} > 0$. The second model is,

Unfortunately, our data do not permit us to identify students who dropped a course.

[12] 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 estimates also test for this hypothesis.

[13] Unfortunately, historic information on which programs require which economics courses is not available.

$$FutureClass_{is} = \begin{cases} 1 \ if \ g_{is}\beta + g_{is}e_s\pi + X_{is}\Gamma + \Phi_s + \varepsilon_{is} < 0 \\ 0 \ if \ g_{is}\beta + g_{is}e_s\pi + X_{is}\Gamma + \Phi_s + \varepsilon_{is} \leq 0 \end{cases}.$$

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 probability of taking subsequent economics courses 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 probability of taking future economics courses 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. Thus, there is no evidence that students take more classes after having more highly rated instructors. Viewed from a revealed preference perspective, this result casts doubt on student evaluations as a measure of teaching quality. For both principles of microeconomics and macroeconomics, students with higher high school class rank and better math ACT or SAT scores are more likely to take subsequent courses while those with higher verbal scores and women are less likely to take additional economics classes. The other controls are not systematically related to the probability of taking additional economics classes.

The estimates in the even-numbered columns of the table show a strong positive 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 and significant for students in intermediate microeconomics courses.

The relationship between grades and the probability of taking future courses only depends on evaluations in intermediate microeconomics. Good students are particularly likely to take subsequent economics classes after taking intermediate microeconomics from a highly-rated instructor, which would bias the relationship between evaluations and learning upward. 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 principles of maroeconomics, which were excluded from the future grade equation from which learning was estimated, 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. 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 3, are reported in Appendix Table 2. Consistent with the previous results, there is a strong relationship between student evaluations and grades, which is unaffected by the inclusion of learning. The learning measure is weakly related to evaluations.

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 students whose prior course is from a less highly-rated instructor. While the estimates of learning based on future grades include fixed effects for future courses, if a disproportionately large number of students from a particular class take courses that are more

19

difficult in the sense of yielding lower average grades for the same amount of initial knowledge, 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 percent of the students in our sample. For students who took an intermediate microeconomics class, we estimate

$$Hard\,Intermediate_{is} = \begin{cases} 1 & if \quad e_s \beta + X_{is} \Gamma + \varepsilon_{is} > 0 \\ 0 & if \quad e_s \beta + X_{is} \Gamma + \varepsilon_{is} \leq 0 \end{cases}$$

using a logit model. 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. There is no relationship between student evaluations in principles of microeconomics or macroeconomics 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 actually less likely to take the more mathematical intermediate microeconomics course, although the coefficient is not statistically significant.) Not surprisingly, students who have higher math ACT or SAT scores are more likely to take the mathematically intensive class.

These estimates indicate that the 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 (in a variety of forms) accounts for the weak relationship between evaluations and learning reported above.

VI. Determinants of Learning and Grades

The preceding 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 the learning measure and how current grades are related to learning and instructor characteristics. These are estimates of equations (4.1) and (4.2). Given that controlling for current grades eliminates the relationship between learning and evaluations, we anticipate that learning and current course grades are positively related.

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. As expected, grades in the current course are positively related to learning, although the relationship is not statistically significant. This finding is consistent with instructors giving higher grades to sections that do better. Not surprisingly students in honors section of the course received higher grades and learned more than other students, and these results are highly statistically significant.

Results for principles of macroeconomics, reported in columns (4) through (6), show that none of the observable instructor characteristics are related to learning, but that women tend to give lower grades. As shown in column (5), learning is positively, but not significantly, related to grades in the current course, but the previous results are robust to controlling for learning.

Results for intermediate microeconomics, reported in columns (7) through (9) show that women tend to assign higher grades, but none of the other instructor characteristics is statistically significantly related to current grades. Students who took intermediate microeconomics from a foreign-born instructor learn less. Otherwise, none of the observable instructor characteristics is statistically significantly related to learning. As before, the current course grades are positively related to learning.

It is noteworthy that Table 2 showed large instructor differences, but that the estimates in Table 7 show no consistent relationship between observable instructor characteristics and learning. 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.

VII. Optimal Evaluation Criteria

This section considers optimal evaluation criteria. Assume that evaluations depend on grades, g_s; learning, l_s; and an unobserved course experience, ξ_s, according to

$$e_s = \phi_g g_s + \phi_l l_s + \xi_s \qquad (3')$$

which is equivalent to (3) without instructor characteristics. If a direct measure of learning and unbiased estimates of $\hat{\phi}_g$ and $\hat{\phi}_l$ are available from (3 / 3'), it is possible to estimate the course experience directly from,

$$\hat{x}_s = e_s - \left(\hat{\phi}_g g_s + \hat{\phi}_l l_s \right).^{[14]} \qquad (5)$$

The $\hat{x}_s$ in (5) are the residuals from (3 / 3'). In other words, our measure of the course experience is evaluations regression-adjusted for grades and learning and (perhaps) instructor characteristics.[15]

With this information, instructors can be evaluated on the course experience they provide and the amount of human capital they produced. With a sense of social priorities, estimates of the course experience, human capital, and grades, administrators can reward instructors based on social welfare.

Because much of the interest is in evaluating instructors, we analyze rankings of instructors for each course.[16] Figure 1.A shows a strong positive relationship between the estimated course experience for instructors and his or her student evaluations for principles of

---

[14] We note that that this procedure attributes all factors that affect evaluations other than grades and learning to the course experience. If, as we have heard mentioned from time to time, instructors give out cookies or the like when administering evaluations, that will be attributed to a better course experience. Thus, in only regression adjusting evaluations, we may be placing too much emphasis on them, but we note that unadjusted evaluations are affected by these biases and our procedure has the advantage of adjusting evaluations for grading leniency.

[15] As noted by the editor the desirability and feasibility of controlling for instructor characteristics is not straightforward. For instance, one might want to adjust for gender to eliminate a gender-bias in evaluations, but it is not clear whether one would want to control for, say, experience.

[16] The course experience for each instructor is the mean of his or her residuals from regression (6) in Table 3, which controls for learning, grades, and section characteristics. The learning produced by each instructor (regression adjusted) is the mean of each instructor's residual of a regression like that in column (6) of Table 3 where learning is the dependent variable and section characteristics are controlled.

microeconomics. The correlation between the two variables (shown in Table 8) is .94 indicating that regression-adjusting evaluations for grades and learning leaves most of the information intact. Panel B plots learning against student evaluations showing virtually no relationship between the amount of learning an instructor produces and his or her student evaluations. Given that learning and evaluations are essentially independent, it is not surprising that many of the instructors who do well on their evaluations do poorly on learning, while many of the instructors who do poorly on their evaluations produce a lot of learning. Thus, student evaluations provide no meaningful information about learning. The table also shows that the correlation between an instructors' course experience and the amount of learning produced is low, .11.

The remaining panels of the table show correlations for principles of macroeconomics and intermediate microeconomics. Again, there is a strong relationship between the imputed course experience and student evaluations, but not between learning and student evaluations. In fact, in these courses, learning is negatively related to evaluations and the course experience, especially for principles of macroeconomics.

The preceding results have two implications for the design of optimal evaluation criteria. First, rankings of instructors based on their course experience are quite similar to rankings of instructors based on raw student evaluations. Nevertheless, we prefer to rank instructors based on the course experience both because it controls for differences in grading leniency and because it eliminates the incentive to inflate grades. Second, rankings of instructors based on learning are markedly different from rankings of instructors based on student evaluations. The optimal assessment of instructors' quality would not be based on raw student evaluations but rather on a weighted average of the course experience and learning variables, where the weights correspond to the relative importance that society assigns to these two objectives.

The optimal assessment of instructor quality is easily implemented. It applies basic regressions to data that are available in machine-readable form at most or all institutions. Four points are worth highlighting. First, constructing optimal evaluations requires one to select the proper weights to be placed on learning relative to the course experience. Second, in order to

properly evaluate the amount of learning, enough time must elapse for students to be observed in subsequent classes. Third, while this measure has the advantage of weighting knowledge in proportion to its importance in future course work, which yields a high level of external validity, the method is most applicable for classes where subsequent coursework will build on the knowledge acquired in the base class. Lastly, the procedure may raise selection concerns, and while we find little evidence for selection in the current context, matters may be different in other contexts. Of course, selection concerns also arise when using conventional student evaluations.

VIII. Conclusions

Our findings show that student evaluations are strongly related to grades and that learning, as measured by future grades, is unrelated to student evaluations once current grades have been controlled. We also provide evidence that evaluations vary with instructor characteristics, the type of section, and composition of the class. We find, for example, that students sometimes give lower evaluations to women and to foreign-born instructors. We do not believe that our results are specific to our institutional setting, and expect our results to be qualitatively similar for higher education generally.

Student evaluations of teaching differ from the ideal construct because they are affected by grade leniency and do not reflect learning produced in a course. Even if student evaluations did not suffer from these two deficiencies, student evaluations would assign proper to weight to learning relative to the course experience only under the highly unlikely condition that students place the same weights on these items as the social planner. Based on our analysis, we propose a procedure for optimally assessing instructor quality. It is simple to use and the underlying data are readily available.

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

26

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. Relationship Between Measures of Teaching Effectiveness – Principles of Microeconomics.
Panel A. Student Evaluations and the Course Experience.
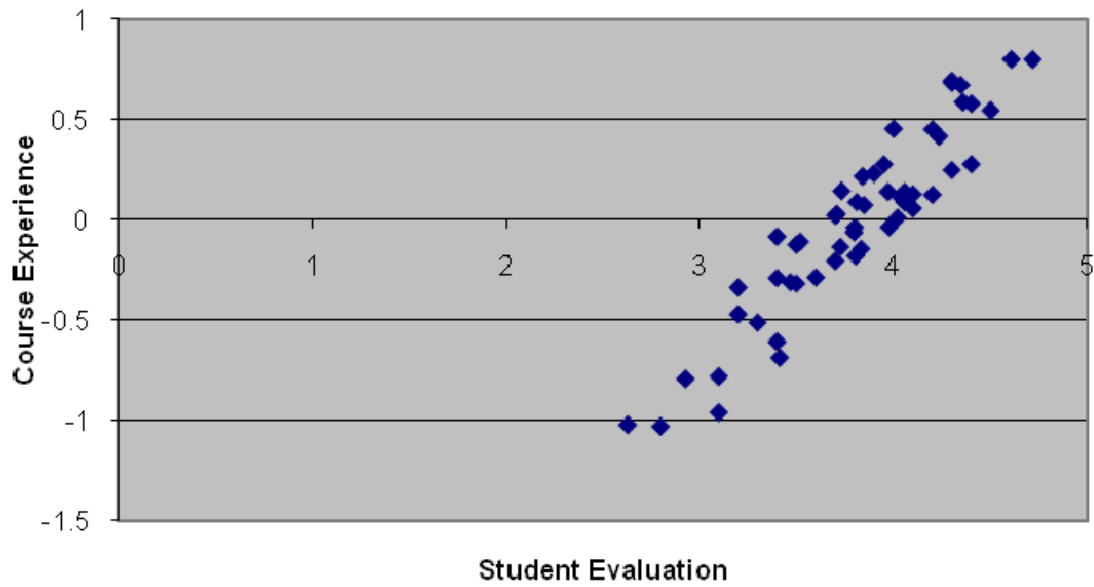
Panel B. Student Evaluations and Learning.
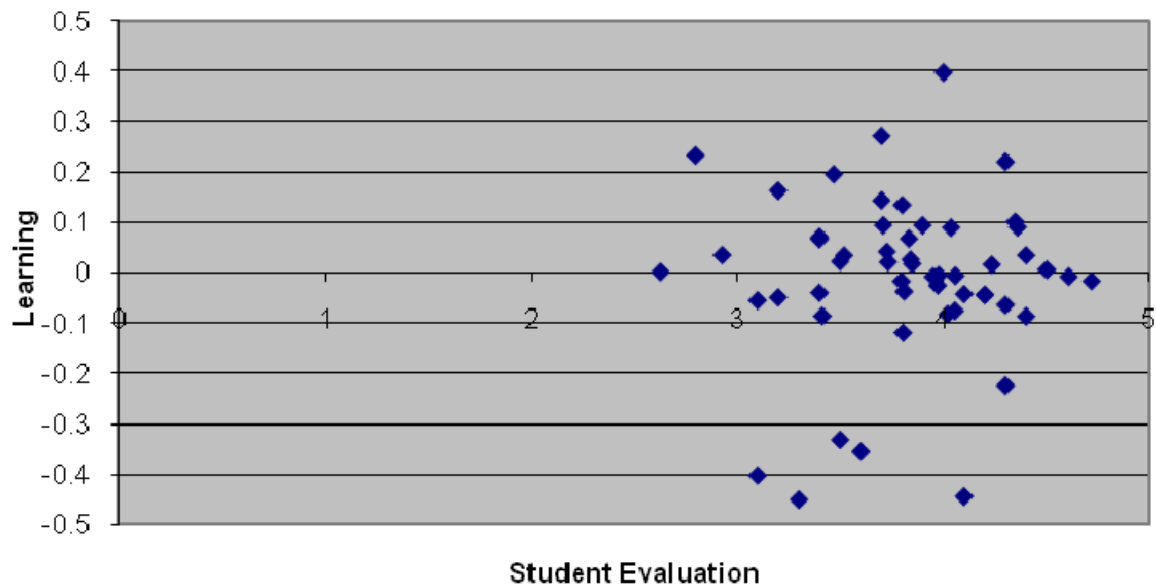
Note. Each point represents an instructor.

Table 1: Sample Characteristics

Variables	Prin. Micro.		Prin. Macro.		Inter. Micro.		Variable Definition
Number of Sections	190		119		86		
Number of Students	26,666		14,729		4,111		
Quality of Teaching	**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 teaching quality
Grades							
Current Course Grade							Current grade, the mean grade the instructor gives
	2.61	(0.30)	2.71	(0.27)	2.72	(0.34)	in the current class (grade leniency)
Learning	2.81	(0.22)	2.84	(0.25)	2.83	(0.24)	Grades the in future Economics courses (regression-adjusted)
Instructor Characteristics							
Instructor: Female	0.32	(0.47)	0.09	(0.28)	0.13	(0.34)	Instructor is female
Instructor: Foreign Born	0.16	(0.37)	0.20	(0.40)	0.31	(0.47)	Instructor is foreign born
Instructor: Lecturer	0.22	(0.41)	0.22	(0.42)	0.18	(0.38)	Instructor is a lecturer
Instructor: Grad. Associate	0.11	(0.32)	0.15	(0.36)	0.22	(0.42)	Instructor is a graduate teaching associate
Instructor: Has Ph.D.	0.81	(0.40)	0.76	(0.43)	0.78	(0.42)	Instructor has a PhD degree
Instructor: Years since Ph.D.	21.58	(13.26)	19.93	(16.79)	10.69	(11.46)	Instructor years since PhD
Instructor: Years at Institution	16.01	(11.96)	16.55	(14.35)	4.78	(5.22)	Instructor years at institution
Student Characteristics							
Student: Female	0.46	(0.50)	0.37	(0.48)	0.30	(0.46)	Student is female
Student: Black	0.08	(0.27)	0.07	(0.26)	0.06	(0.24)	Student is black
Student: Hispanic	0.02	(0.14)	0.02	(0.13)	0.01	(0.12)	Student is hispanic
Student: HS class rank	55.76	(34.39)	54.95	(34.28)	42.59	(38.46)	Student High School rank
Student: ACT English	22.03	(4.41)	22.05	(4.40)	22.25	(4.81)	Student ACT English
Student: Math	23.13	(4.39)	23.45	(4.26)	23.80	(4.47)	Student ACT math
Student: Reasoning	22.86	(3.97)	23.03	(3.95)	23.43	(4.24)	Student ACT reasoning
Student: SAT Verbal	52.04	(8.94)	51.58	(9.21)	51.14	(10.84)	Student SAT verbal
Student: SAT Math	55.27	(9.09)	55.87	(8.88)	56.79	(9.91)	Student SAT math
Class Characteristics							
Multi-Section Class	0.68	(0.47)	0.66	(0.48)	N/A	N/A	Class is a large lecture with multiple small recitation sections
Honors Class	0.06	(0.24)	0.07	(0.26)	N/A	N/A	Class is a honors class
Night Class	0.14	(0.35)	0.13	(0.34)	0.26	(0.44)	Class is a night class
Calculus Class	N/A	N/A	N/A	N/A	0.12	(0.33)	Course is calculus based
Response Rate	0.54	(0.21)	0.54	(0.19)	0.62	(0.14)	Response rate on SET survey

Note: Standard deviations in parentheses.

Table 2: Tests of Grades in Future Sections as Measures of Base-Section Learning.

	(1)	(2)	(3)
Base Course:	Micro-Principles	Macro-Principles	Intermediate Micro
Standard Deviation in Learning across Sections			
Standard deviation	0.146	0.192	0.152
Relative to variation in future grades across sections	0.672	0.831	0.458
Test for Joint Significance of Base-Section Dummy Variables as Determinants of Learning in Equation (2)			
F-Statistic	2.99	3.56	3.56
Degrees of Freedom (Numerator, Denominator)	(327, 34204)	(297, 20866)	(150, 9106)
P-Value	<.0001	<.0001	<.0001
Base-Section Instructors as Determinants of Learning			
F-Statistic for Joint Significance of Instructor-Effects	1.7	3.51	1.52
Degrees of Freedom (Numerator, Denominator)	(52, 137)	(31, 87)	(29, 56)
P-Value	0.008	<.0001	0.0904
R^2 of Instructor Effects	0.392	0.555	0.44

31

Table 3: Determinants of SET Overall Rating – Principle Microeconomics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Current Course Grade	0.44***		0.40***		0.42***	0.37***	0.39***	0.44***	0.48***
	(0.10)		(0.10)		(0.10)	(0.11)	(0.11)	(0.14)	(0.14)
Learning		0.25**	0.14		0.13	0.09	-0.05	-0.19	-0.17
		(0.11)	(0.11)		(0.11)	(0.11)	(0.14)	(0.18)	(0.19)
Lag of Current Course Grade									-0.00
									(0.12)
Instructor: Female				-0.27*	-0.28*		-0.26*	-0.10	-0.14
				(0.16)	(0.15)		(0.15)	(0.23)	(0.19)
Instructor: Foreign Born				-0.16	-0.24		-0.23	-0.48*	-0.53**
				(0.18)	(0.17)		(0.17)	(0.26)	(0.21)
Instructor: Lecturer				0.04	0.07		0.04	-0.09	-0.07
				(0.28)	(0.26)		(0.28)	(0.37)	(0.30)
Instructor: Graduate Student				0.05	0.02		0.06	0.12	0.06
				(0.55)	(0.52)		(0.56)	(0.71)	(0.58)
Instructor: Has Ph.D.				-0.01	-0.06		-0.13	-0.28	-0.28
				(0.57)	(0.54)		(0.57)	(0.77)	(0.62)
Instructor: Years since Ph.D.				0.00	0.00		0.01	0.01	0.00
				(0.01)	(0.01)		(0.01)	(0.01)	(0.01)
Instructor: Years at Institution				0.00	0.00		0.00	0.00	0.00
				(0.01)	(0.01)		(0.01)	(0.01)	(0.01)
Multi-Section Class						0.05	-0.08	-0.11	-0.07
						(0.12)	(0.15)	(0.17)	(0.18)
Honors Class						0.22**	0.21**	0.27**	0.27**
						(0.09)	(0.11)	(0.12)	(0.13)
Night Class						0.15**	0.16**	0.19**	0.18*
						(0.07)	(0.07)	(0.09)	(0.10)
Response Rate							0.03	-0.06	-0.08
							(0.20)	(0.25)	(0.26)
Constant	4.02***	3.77***	3.97***	3.93***	4.20***	3.87***	4.02***	4.12***	4.27***
	(0.08)	(0.07)	(0.08)	(0.58)	(0.55)	(0.09)	(0.63)	(0.89)	(0.77)
Includes Year Fixed Effects							Yes	Yes	Yes
Includes Student Characteristics as Controls in First Stage	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	190	190	190	190	190	190	190	133	133
R-square	0.1283	0.0224	0.1379	0.0057	0.2036	0.2319	0.2772	0.4865	0.5335

Note: Standard errors in parentheses. Estimates include random effects for instructors. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4: Determinants of SET Overall Rating – Principle Microeconomics – Robustness Checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Section-Level Estimates, Learning Measured by Performance in Intermediate Microeconomics									
Current Course Grade	0.37*** (0.10)		0.38*** (0.10)		0.38*** (0.10)	0.34*** (0.11)	0.35*** (0.11)	0.46*** (0.15)	0.46*** (0.15)
Learning (Measured by Performance in Intermediate Microeconomics)		0.02 (0.05)	-0.01 (0.05)		-0.01 (0.05)	-0.02 (0.05)	-0.03 (0.06)	-0.09 (0.09)	-0.11 (0.09)
Lag of Current Course Grade									0.20 (0.14)
Observations	175	175	175	175	175	175	175	122	122
R-square	0.1162	0.0014	0.1156	0.0011	0.1181	0.2165	0.2182	0.4803	0.5034
Instructor-Level Estimates, Learning Measured by Performance in All Future Economics Classes									
Current Course Grade	0.56** (0.25)		0.56** (0.25)		0.64** (0.26)	0.75** (0.29)	0.64* (0.33)		
Learning (Measured by Performance in All Future Classes)		-0.06 (0.30)	0.02 (0.29)		-0.00 (0.30)	0.03 (0.30)	0.08 (0.33)		
Observations	53	53	53	53	53	53	53	53	53
R-square	0.093	0.001	0.093	0.141	0.248	0.138	0.302	0.381	0.508
Includes Year Fixed Effects							Yes	Yes	Yes
Includes Instructor Characteristics				Yes	Yes		Yes	Yes	Yes
Includes Section Characteristics						Yes	Yes	Yes	Yes
Includes Student Characteristics as Controls in First Stage	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses. Estimates in the top panel include random effects for instructors. * Significant at 10%; ** significant at 5%; *** significant at 1%. For the instructor-level estimates, we construct the weighted mean of student evaluations and of our measures of grades and learning for each section taught by an instructor estimated in equations (1) and (2). In these calculations, each section was weighted by the number of students in that section. Equation (3) is then estimated at the instructor level with these instructor-level means as the independent and dependent variables. In the specifications that include section and instructor characteristics, the means of those variables are included.

Table 5: Determinants of the Taking Subsequent Economics Courses

	Principles of Micro		Principles of Macro		Intermediate Micro	
	(1)	(2)	(3)	(4)	(5)	(6)
SET Overall Rating	0.06		0.07		-0.01	
	(0.06)		(0.10)		(0.08)	
Course Grade		0.54***		1.02***		-0.43**
		(0.14)		(0.21)		(0.20)
SET Overall Rating * Course Grade		-0.00		-0.09		0.18***
		(0.03)		(0.06)		(0.06)
Student: Female	-0.51***	-0.49***	-0.20***	-0.16***	-0.14	-0.14
	(0.03)	(0.03)	(0.04)	(0.04)	(0.09)	(0.10)
Student: Black	-0.03	0.14***	-0.26***	-0.05	-0.20	-0.17
	(0.04)	(0.05)	(0.08)	(0.08)	(0.17)	(0.17)
Student: Hispanic	0.14	0.20*	-0.15	-0.04	-0.31	-0.34
	(0.11)	(0.12)	(0.14)	(0.14)	(0.35)	(0.35)
Student: High School Rank	0.00***	-0.00	0.00***	-0.00	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Student: ACT English	-0.03***	-0.04***	-0.01	-0.02***	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Student: ACT Math	0.04***	0.02***	0.05***	0.03***	0.02	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Student: ACT Reasoning	0.00	0.00	-0.01	-0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Student: SAT Verbal	-0.01	-0.01	-0.02***	-0.02***	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Student: SAT Math	0.02***	0.01**	0.03***	0.02***	0.01	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Multi-Section Class	0.27**	0.08	0.05	1.27***		
	(0.10)	(0.10)	(0.14)	(0.22)		
Honors Class	0.46***	1.81***	0.15	-0.96*		
	(0.12)	(0.14)	(0.14)	(0.49)		
Night Class	-0.02	-0.09	-0.19**	-0.06	-0.50***	-1.12***
	(0.09)	(0.18)	(0.09)	(0.12)	(0.15)	(0.26)
Calculus Class					0.55***	-1.16***
					(0.17)	(0.19)
Instructor: Female	-0.10*		0.16		-0.25	
	(0.06)		(0.11)		(0.16)	
Instructor: Foreign Born	-0.12		0.14		-0.17	
	(0.09)		(0.14)		(0.15)	
Instructor: Lecturer	0.09		-0.39**		0.58	
	(0.10)		(0.18)		(0.39)	
Instructor: Graduate Student	0.19		-0.75***			
	(0.24)		(0.20)			
Instructor: Has Ph.D.	0.25		-0.88***		0.12	
	(0.24)		(0.19)		(0.17)	
Instructor: Years since Ph.D.	-0.01*		0.03***		-0.01	
	(0.01)		(0.01)		(0.01)	
Instructor: Years at Institution	0.00		-0.03***		0.01	
	(0.01)		(0.01)		(0.02)	
Constant	-313.73	-347.94	311.92	1,757.28	-686.29	-393.95
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Includes Section Fixed Effects		Yes		Yes		Yes
Controls for Inclusion of High School Rank & SAT/ ACT	Yes	Yes	Yes	Yes	Yes	Yes
Observations	26575	26335	11164	10967	3989	3912

Note: Estimates from Logit models. Standard errors, which are clustered at the section-level, in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%

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

	(1) Micro-Principles	(2) Macro-Principles
SEI Overall Rating	0.15	-0.18
	(0.20)	(0.14)
Instructor: Female	0.21	-0.47*
	(0.16)	(0.25)
Instructor: Foreign Born	0.05	-0.01
	(0.26)	(0.24)
Instructor: Lecturer	-0.17	-0.00
	(0.36)	(0.33)
Instructor: Graduate Student	-0.01	0.39
	(0.78)	(0.44)
Instructor: Has Ph.D.	0.23	-0.08
	(0.68)	(0.44)
Instructor: Years since Ph.D.	0.00	-0.01
	(0.01)	(0.02)
Instructor: Years at Institution	0.00	0.01
	(0.01)	(0.02)
Student: Female	-0.07	-0.27*
	(0.15)	(0.16)
Student: Black	0.66*	0.51*
	(0.34)	(0.30)
Student: Hispanic	0.49	0.32
	(0.38)	(0.45)
Student: High School Rank	0.00	0.00
	(0.01)	(0.00)
Student: ACT English	0.00	-0.01
	(0.02)	(0.03)
Student: ACT Math	0.16***	0.13***
	(0.02)	(0.03)
Student: ACT Reasoning	0.03	0.02
	(0.03)	(0.03)
Student: SAT Verbal	-0.01	0.00
	(0.02)	(0.02)
Student: SAT Math	0.08***	0.06***
	(0.02)	(0.02)
Multi-Section Class	-0.22	-0.51*
	(0.37)	(0.30)
Honors Class	0.43	0.80***
	(0.37)	(0.27)
Night Class	0.08	-0.83***
	(0.33)	(0.24)
Includes Year Fixed Effects	Yes	Yes
Controls for Inclusion of High School Ranks and SAT / ACT	Yes	Yes
Observations	2184	2511

Note: Estimates from logit models. Standard errors, which are clustered at the section-level, in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 7: Determinants of Course Grades and Learning

	(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		0.08			0.15			0.25*	
		(0.10)			(0.10)			(0.13)	
Instructor: Female	0.06	0.05	0.06	-0.22*	-0.24*	0.07	0.26**	0.27**	-0.07
	(0.08)	(0.08)	(0.06)	(0.13)	(0.14)	(0.08)	(0.13)	(0.13)	(0.08)
Instructor: Foreign Born	0.11	0.11	0.04	0.07	0.09	-0.08	-0.01	0.04	-0.14**
	(0.09)	(0.09)	(0.06)	(0.12)	(0.13)	(0.07)	(0.12)	(0.12)	(0.07)
Instructor: Lecturer	-0.01	-0.02	0.13	0.07	0.08	-0.00	-0.51	-0.51	0.07
	(0.15)	(0.15)	(0.10)	(0.19)	(0.20)	(0.11)	(0.33)	(0.35)	(0.19)
Instructor: Grad. Associate	0.20	0.18	0.20	0.30	0.30	0.04			
	(0.29)	(0.30)	(0.20)	(0.27)	(0.28)	(0.14)			
Instructor: Has Ph.D.	0.10	0.08	0.10	0.56*	0.55*	0.02	0.12	0.12	-0.03
	(0.29)	(0.30)	(0.20)	(0.29)	(0.30)	(0.16)	(0.15)	(0.15)	(0.09)
Instructor: Years since Ph.D.	-0.00	-0.00	0.00	-0.02	-0.02	0.00	0.00	0.00	-0.00
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Instructor: Years at Institution	-0.00	-0.00	0.00	0.01	0.01	-0.00	-0.01	-0.01	0.01
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
Multi-Section Class	0.13	0.14	-0.05	-0.05	-0.02	-0.13			
	(0.10)	(0.10)	(0.08)	(0.13)	(0.13)	(0.10)			
Honors Class	0.25***	0.24***	0.25***	0.21***	0.17**	0.29***			
	(0.06)	(0.07)	(0.05)	(0.07)	(0.08)	(0.07)			
Night Class	-0.07	-0.07	0.01	-0.01	-0.01	-0.04	0.05	0.07	-0.01
	(0.05)	(0.05)	(0.04)	(0.07)	(0.07)	(0.06)	(0.08)	(0.07)	(0.07)
Calculus-Based Class							0.01	-0.03	0.19***
							(0.08)	(0.08)	(0.07)
Constant	-0.91***	-0.89**	-0.18	0.40	0.45	-0.28*	-1.00***	-0.78***	-0.94***
	(0.34)	(0.35)	(0.24)	(0.30)	(0.31)	(0.17)	(0.15)	(0.19)	(0.12)
Number of Sections	190	190	190	119	119	119	86	86	86
R-Square	0.2001	0.1971	0.4908	0.3695	0.3744	0.6094	0.4488	0.4783	0.2916

Note: Standard errors in parentheses. Estimates include random effects for instructors. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 8: Correlations between Measures of Teacher Quality

Principle Microeconomics	Student Evaluation	Course Experience	Learning (Regression Adjusted)
Student Evaluation	1		
Course Experience	0.9373	1	
Learning (Regression Adjusted)	0.0404	0.1076	1
Principle Macroeconomics			
Student Evaluation	1		
Course Experience	0.9590	1	
Learning (Regression Adjusted)	-0.2629	-0.2459	1
Intermediate Microeconomics			
Student Evaluation	1		
Course Experience	0.8169	1	
Learning (Regression Adjusted)	-0.0943	-0.0890	1

Note: The course experience is the mean of each instructor's residual of regression (6) in Table 3, which controls for learning, grades, and section characteristics. Learning (regression adjusted) is the mean of each instructor's residual of a regression like that in column (6) of Table 3 where learning is the dependent variable and section characteristics are controlled.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Principles of Macroeconomics									
Current Course Grade	0.60***		0.55***		0.56***	0.56***	0.47***	1.01***	0.94***
	(0.15)		(0.16)		(0.17)	(0.17)	(0.18)	(0.22)	(0.22)
Learning		0.29**	0.13		0.17	0.16	0.10	0.04	0.02
		(0.14)	(0.14)		(0.15)	(0.15)	(0.18)	(0.22)	(0.21)
Lag of Current Course Grade									0.53**
									(0.23)
Instructor: Female				-0.60**	-0.46*		-0.40	0.38	0.53*
				(0.30)	(0.27)		(0.26)	(0.28)	(0.28)
Instructor: Foreign Born				-0.19	-0.28		-0.38	-0.17	-0.18
				(0.27)	(0.24)		(0.24)	(0.20)	(0.19)
Instructor: Lecturer				0.22	0.00		-0.24	-0.90***	-0.96***
				(0.41)	(0.37)		(0.37)	(0.31)	(0.30)
Instructor: Graduate Student				-0.00	-0.33		-0.56	-1.41***	-1.59***
				(0.63)	(0.57)		(0.53)	(0.40)	(0.39)
Instructor: Has Ph.D.				0.01	-0.31		-0.23	-1.81***	-2.24***
				(0.67)	(0.61)		(0.57)	(0.50)	(0.52)
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	-0.08***	-0.09***
				(0.02)	(0.02)		(0.02)	(0.02)	(0.02)
Number of Sections	119	119	119	119	119	119	119	85	85
R-Square	0.1262	0.0272	0.1266	0.2831	0.3991	0.1219	0.4568	0.6424	0.6715
Intermediate Microeconomics									
Current Course Grade	0.62***		0.63***		0.68***	0.57***	0.52**	0.98**	1.05**
	(0.22)		(0.23)		(0.25)	(0.20)	(0.23)	(0.39)	(0.41)
Learning		0.04	-0.05		-0.02	-0.24	-0.30	-0.56	-0.52
		(0.27)	(0.27)		(0.28)	(0.23)	(0.23)	(0.40)	(0.41)
Lag of Current Course Grade									-0.33
									(0.45)
Instructor: Female				0.21	0.07		0.26	-0.85**	-0.75*
				(0.27)	(0.26)		(0.31)	(0.38)	(0.41)
Instructor: Foreign Born				-0.06	-0.16		-0.24	-0.83***	-0.83***
				(0.24)	(0.23)		(0.29)	(0.22)	(0.22)
Instructor: Lecturer				0.16	0.20		0.40	-0.03	-0.02
				(0.68)	(0.65)		(0.84)	(0.60)	(0.60)
Instructor: Has Ph.D.				-0.39	-0.42		-0.63*	-0.29	-0.34
				(0.29)	(0.27)		(0.36)	(0.35)	(0.36)
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	0.01	0.01
				(0.03)	(0.03)		(0.04)	(0.04)	(0.04)
Number of Sections	86	86	86	86	86	86	86	53	53
R-Square	0.0801	0.0004	0.0809	0.0344	0.1644	0.1634	0.3665	0.6160	0.6224
Includes Student Characteristics							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. Estimates include random effects for instructors. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table 2: Determinants of SET Overall Rating – Principle Microeconomics – Selection Corrected Learning Measure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Current Course Grade	0.35***		0.34***		0.34***	0.28**	0.24**	0.52***	0.39**
	(0.10)		(0.10)		(0.11)	(0.11)	(0.12)	(0.17)	(0.18)
Learning (Selection Corrected)		0.10	0.06		0.08	0.05	0.07	-0.10	-0.09
		(0.09)	(0.09)		(0.09)	(0.09)	(0.10)	(0.17)	(0.17)
Lag of Current Course Grade									0.29*
									(0.15)
Instructor: Female				-0.37***	-0.36***		-0.31**	-0.30**	-0.29**
				(0.13)	(0.13)		(0.13)	(0.14)	(0.13)
Instructor: Foreign Born				-0.29**	-0.30**		-0.34**	-0.87***	-0.81***
				(0.14)	(0.14)		(0.15)	(0.14)	(0.14)
Instructor: Lecturer				-0.18	-0.08		-0.05	-0.10	0.04
				(0.22)	(0.22)		(0.23)	(0.21)	(0.22)
Instructor: Graduate Student				-0.51	-0.41		-0.14	-0.40	-0.19
				(0.44)	(0.44)		(0.47)	(0.43)	(0.44)
Instructor: Has Ph.D.				-0.32	-0.28		-0.06	-0.32	-0.14
				(0.46)	(0.46)		(0.47)	(0.42)	(0.42)
Instructor: Years since Ph.D.				-0.01	-0.01		-0.01	-0.02**	-0.02**
				(0.01)	(0.01)		(0.01)	(0.01)	(0.01)
Instructor: Years at Institution				-0.01	-0.00		-0.00	-0.01	-0.00
				(0.01)	(0.01)		(0.01)	(0.01)	(0.01)
Multi-Section Class						0.08	0.05	0.17	0.25
						(0.12)	(0.15)	(0.18)	(0.18)
Honors Class						0.28***	0.22**	0.49***	0.55***
						(0.09)	(0.11)	(0.16)	(0.16)
Night Class						0.15*	0.14*	0.24**	0.28**
						(0.08)	(0.08)	(0.12)	(0.12)
Response Rate							0.21	-0.21	-0.12
							(0.21)	(0.30)	(0.30)
Includes Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Sections	183	183	183	183	183	183	183	183	129
R-Square	0.2156	0.1245	0.2244	0.2797	0.4116	0.3441	0.4930	0.7390	0.7490

Note: Standard errors in parentheses. Estimates include random effects for instructors. * 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.