Creating a True Measure of Instructor Quality and Student Learning at the University of Oregon

Ethan Uecker
&
Devin Olson

Presented to the Department of Economics at the University of Oregon, in partial fulfillment of the requirements for honors in Economics

June 9, 2017

Under the supervision of Bill Harbaugh University of Oregon

Abstract:

In this paper we develop two objective quantitative measures of teaching effectiveness and estimate them for instructors at the University of Oregon. The first measure shows what percentage of an instructor's students go on to take a second course in the subject. The second shows how an instructor's students do in a second course in the subject relative to students who take the first class from other instructors. We show a way to measure students' actual value added by an instructor by creating a proxy measure of deep learning as an alternative to student evaluations of teaching, which are not an effective measure.

Table of Contents:

Introduction:

The University of Oregon, like most universities, uses student evaluations of teaching (SETs) to evaluate the quality of teaching and provide instructors with student input to improve their teaching. The information provided by these evaluations are used to make decisions within departments and the university as a whole about pay, instructor awards, tenure and promotion. Although important in the university system, recent studies on the validity of these student evaluations as measures of teaching quality have shown that there is no significant correlation between student evaluations of teaching, and teaching quality. It is sometimes found that there is a positive correlation between the evaluations and grades in the current course, but a negative correlation between the evaluations and grades future courses, suggesting that high scores on evaluations may be correlated with instructors that may be "easy-graders" but not preparing students for future classes.

Another issue with SETs is the empirical evidence of biases in the scores according to the gender, race or even the physical appearance of the instructor. It is difficult to determine the true effects of these biases when the measure of "teaching quality" may not be an accurate measure itself.

Quality of teaching isn't the only factor that should be looked at when evaluating a teacher. Another important quality of an instructor to the university is how well the instructor encourages students to continue with their education. This is especially important for departments with decreasing enrollment. A teacher that can effectively improve students' abilities in a subject in addition to encouraging them to continue with their study within that subject is a teacher that departments and universities should be seeking.

This paper seeks to find a more accurate measure of teaching quality by creating more objective measures of determining teaching ability. Students receiving a good education from quality teachers should have noticeable gains in their achievement levels from class to class, relative to students taking initial courses from worse instructors. We hope universities will revamp their evaluation methods by possibly dropping the biased and inaccurate SET score method and adopting this more objective measure of teaching quality.

Our results indicate that by looking at both an instructor's continuation rate, as well as their value added from one class to the next, we can create a quantitative measure for teacher quality. Furthermore, these measures control for student characteristics and across-department variations, which SETs typically do not.

Literature Review

This paper builds on recent research that has called into the question of validity of SETs as an accurate measure of teaching quality as well as the concerns with biases within SETs.

The research done at the US Air Force Academy by Carrell and West (Journal of Political Economy, 2010) is perfect for exploring the relationship of teacher quality on future student performance. They again measured this both by looking at grades in common exams and future class performance. At the Air Force academy, all tests are standardized across classes and professors. Student placement is random in regard to their teacher. This makes it easy to control for differences in grading leniency.

Teacher quality measurements included their academic rank, gender, years of experience, and education level. They find that "less experienced and less qualified professors produce students who perform significantly better in the contemporaneous course being taught, whereas more experienced and highly qualified professors produce students who perform better in the follow-on related curriculum." This is perhaps accredited to the fact that newer teachers adhere more strictly to teaching guidelines, or care more about their evaluation scores, so they simply teach their students what they need to know for the tests. In addition, their data showed that students rewarded these kinds of professors with higher scores, as again, a higher grade is positively correlated with higher SET scores. This paper shows that subsequent performance is a better indicator of teaching quality than SETs.

In their paper titled *Student evaluations of teaching (mostly) do not measure teaching effectiveness*, Boring, Ottoboni and Stark explore different kinds of SET bias. They find that SET scores are more sensitive to students' gender bias and grade expectations than they are to teacher effectiveness. They also show that gender biases alone can be large enough to cause more effective instructors to actually receive lower SET than less effective instructors.

They controlled for randomization using two separate experiments; the first is the French Natural Experiment, the second is a randomized US experiment. In the French university system students have to take 6 mandatory courses their freshman year and are randomly assigned their instructors. In addition, they all have the same anonymously graded final exam. Using this exam, they looked at the correlation between mean SET scores and mean final exam scores. They expected them to be positively correlated if SET is a true measurement of teaching quality.

The US experiment allows us to better control for differences in teaching style across instructors. They collected SET data from an online course in which a group of students were assigned to 4 discussion groups, each taught by one male and one female TA. The biases were revealed by differences in their ratings when the TA genders were given to students. So even though the students never actually saw their teachers, the one they perceived to be male consistently got higher SET scores.

Our paper is perhaps closest to Figlio, Schapiro and Soter (2015), which measures the difference in teaching quality between tenured and non-tenured instructors. Their conclusion is that students learn more from contingent instructors than tenured instructors. They use data from first-year students at Northwestern University. They control for student-level fixed effects and next-class-taken-fixed effects to control for other factors that could impact the learning of to the students in a given class. Their simplest model has independent variable L(a dummy variable for a contingent instructor) and dependent variable G(GP of student in the next class taken in the same subject). They use this model to get a measure of "value added" by the staff members. This "value added" is the basis for value that we will be using. They also include a variable to only compare students who took the same instructor and class combination for the first and second courses within a subject. They do this to avoid having variation between the instructors of the second courses that would make the measure of "value added" inaccurate. They also looked at the probability a student would take another class within the subject dependent on their instructor type in the first course (tenured or non-tenured). Their results showed that the worst value-adding tenured professors drove the consistent results that tenured instructors as a whole did worse with the "value added" measure than non-tenured instructors as a whole.

Uttl, White, Gonzalez conducted a meta-analysis of the teaching effectiveness of university instructors. They chose studies with the following restrictions: they dealt with the connection between evaluations and learning in a university or college setting, involved multiple sections of the same course, evaluations and measures of learning had to be consistent across sections within the study, measures of learning measured actual learning and not the student's perspective of learning, evaluation and measures of learning correlations were calculated using section means and not individual student scores. Their conclusion was that there is no significant correlation between evaluations and learning. The evaluations instead are good indications of students satisfaction with the course, regardless of how much they actually learn.

Finally, Ancell and Wu (2017) took a unique approach when looking at the relationship between numerical SET scores and student success, and at potential biases in the SET's based on gender and race/ethnicity. They first examined how evaluation scores were a function of instructor characteristics, course characteristics, and class GPA. They found that academic rank, class size, and instructor gender were all subject to biases in SETs. In addition, they concluded that the effect of grades on SET scores was large and consistent, but the root cause of this effect was difficult to pinpoint. Ultimately, if course evaluation scores are a true unbiased "valid measure of teaching quality, then changes in instructor characteristics shouldn't cause any variability across evaluation scores."

Their second model is similar but compares the relationship of course evaluation scores to teaching quality. Their measure of teaching quality was average student performance in a future course in a sequence. If students truly learn more from better teachers, then they should perform better in subsequent classes. To control for differences in teaching methods and grading leniency, they normalized class scores by observing relative class rank. Specifically, they measured student achievement as a function of student, instructor, and class characteristics, as well as course evaluation questions. They used data from the School of Journalism and Communications and the Lundquist College of Business for their research.

Results from these two methods were as follows. Female instructors receive relatively worse scores than their male counterparts. In addition, they found career instructors had slightly higher SET scores. This seems counter-intuitive, as tenured professors should have superior teaching quality, but it could be due to them being driven more by the performance of students on tests than them actually learning material. As expected, a higher average GPA in the class leads to higher corresponding SET scores. As for future student achievement, having a tenured professor, a male professor, and a smaller class all had positive effects on subsequent classes. The tenured professors had lower SET scores, but their students had higher future achievement. This is more evidence that the SETs are not correlated with student learning.

Although a good starting point, their research and findings had several limitations. They only used data from the Business and Journalism school. Therefore, the don't get a wide range of subjects, or any from STEM subjects such as math or and of the sciences. In addition, they don't fully develop an alternative measure, nor do they look at continuation and its possible selection effect on value added. There paper also differs from ours in using the first and second courses in specific sequences. We use the more comprehensive and arguably more general method of looking at the first and second courses taken in a given subject - whether these were in a defined sequence or not.

Data:

We will be using three data sets that begin in fall term 2000 and end in winter 2018: student transcript data, faculty information and registrar data provided by the University of Oregon. The transcript and registrar data include data on each class a student took, the grading option they chose, and their grade in the class. However, we had to piece the data we had together, and every student is given a unique identification number, so we don't have names. In addition, it has information on student demographics such as their gender, ethnicity, and birth year. We also have some measures of their past performance as students such as their high school GPA, ACT and SAT scores, and how many credits they had previously earned prior to their enrollment at the University of Oregon. This data also specifies whether the student transferred from another university, came to the university directly after completing high school, or if they earned a GED. We also know in which term they first enrolled in the university and in which term they graduated, if they have done so. The data also included whether or not the student received a Pell grant, a federal grant provided to students with sufficient financial need.

We constructed several new variables from this data set. In our faculty data set we have instructor demographics such as their names and race. We used a package through R that uses the first names of the professors to predict their genders, allowing us to include instructor gender in our data. R is an open-source statistical software program, and is often used for research such as this (Rstudio.com). It has several programs that can predict certain characteristics or attributes (i.e. gender), which is what we used in this case. We did not have access to names for the graduate student employees (GE) and therefore could not obtain genders for them. We also know instructor respective rank at the university which we can use to determine if they are tenured or non-tenured faculty. We also have instructor salary (which is public information) and what classes they have taught. We then merged these data sets into one large comprehensive data set.

We converted the letter grade received and a few other variables into numeric variables. We converted their letter grade into corresponding number values (i.e. a 4.0 is equivalent to an A). We also created grade points for classes that were taken as a Pass/No Pass option. We set "no pass" to a GP of 0 and "pass" to a GP of 2. We then used this to find quality points earned by multiplying the grade point (GP) by the amount of credits the class was worth. We also added up credits taken per term by a student to get term credits per student for each term and calculated the quality points for the term. Then we used this data to calculate their term GPA.

UO identifies courses with Course Reference Numbers (CRNs) that are reused every few years, so we created a unique identifier for each course using the CRN and the term the class took place in. By using this identifier and looking at how many students were enrolled in a given class in a term, we created a variable for class size.

We also had to create several dummy variables. For student ethnicity we were given a number for each student which corresponded to a different race. Once we decoded each student's race, we created a dummy variable for each possibility (nine in total). We also created a dummy variable for whether or not the student was male. For instructor data we created a dummy variable for whether or not the instructor was male and one dummy variable for each instructor type.

Once we had the initial data sets combined and variables created, we created the data set that we used for our regressions. In order to measure teaching quality, we needed a way to look at future student performance. First, we took all the first classes students took in a particular subject. This isn't necessarily in their first year, as they could have taken their first course in Economics for example in their second year, then continued with Economics classes after that. This is a potential limitation in our paper, as we don't have a perfect control for what year they began taking classes in. Then we created a second data set, which consisted of students who took a second class in that subject. Finally, we merged the two smaller datasets by matching up each student and instructor, and created a dummy variable called "continued" if they had taken a second class in the subject. We did this for the 20 most popular subjects, measured by how many students enroll in the correlating major. It is important to note that most first and second-class pairs will be lower level classes. This is because we take the very first class and the second-class students take in a subject and pair those together. This may limit our model's overall usefulness, but we could extend the research into sequences. The university's course catalog has every major, and every major requirement. For some more rigorous and defined majors (i.e. chemistry, physics, math, etc.) there are very specific sequences they must take, one class right after another. This may provide us with new insight and be able to control for the aforementioned weakness of our method.

We used this data set to create the two main variables our research was concerned with, the "Ratio Continued" and "Value added" Measures. The goal of the Value-Added measure is to compare how an instructor's students do in future courses in the same subject relative to their peers who took the first course from a different instructor. If an instructor teaches students that go on to do better than their peers in future classes, controlling for their initial grade in that subject, it is likely that the teacher is providing the opportunity for those students to achieve deep learning that his helpful in future courses. We must use relative grades rather than raw grades to control for grading leniency between classes. For example, a student who gets a 3.0 in a class where the average is a 2.5 is doing relatively better than a student who gets a 3.5 and the class average is also a 3.5. To calculate this measure, we first calculated the GPA of each class as a whole. Then we created a measure for each student's GP relative to the other students in their class by taking each student's GP in the class and dividing it by the class average. If the student continued with the subject and took another class within the subject in the future then we subtracted the student's relative GP in the first class from the student's relative GP in the second class to get the value added measure for each student.

One flaw in this Value-Added measure is it doesn't account for selection effects. If a teacher is terrible at teaching and a tough grader, then it is likely that very few students will continue with future courses in the subject, leaving only the most intelligent students that were able to do well despite their instructor's inability to teach. Because this teacher only encourages the brightest students to continue then it is likely that these students will do well in future classes relative to their peers, which would give their first professor a high score in our value-added measure. Similarly, an excellent teacher would be able to effectively help struggling students and encourage many students to continue with the subject despite their possible difficulty with the course material. These students that continue despite their struggles are likely to not succeed in future courses without excellent instruction, causing the initial instructor's ability to be underestimated by the value-added measure.

To account for this selection effect, we created a "Ratio Continued" variable to control for how many students continue with a subject when we look at the value-added measure. This Ratio Continued variable is the ratio of students from a given class that continue with the subject. This variable allows us to control for "grading leniency" and other factors that may be taken into account when students decide whether they will take another course in the same subject after being taught by a specific teacher in our Value-Added regressions.

From this paired data set, we dropped observations that were more than two terms apart. We did this because we wanted to test deep learning, but not learning that was so far removed from the first class that the effects of the first instructor would be diminished. We also dropped students who were graduating, as they obviously would not be able to have a second class to match the first one if they didn't continue to attend the university anymore. However, we did keep students who dropped out of the university after the first class, as their instructor could have been an influence on their decision to not continue at the University of Oregon.

Our original data set, with every student and every class they have taken since 2000, has 3,404,056 observations. After dropping all observations because of factors mentioned above, as well as only keeping the most popular 20 subjects, and pairing students with their first and second courses taken in a subject, we are left with a total observation count of 489,989.

For this research, we used the 20 most popular majors here at the University of Oregon. The following table shows these majors and the average and standard deviation for Ratio Continued:

SUBJ	MEAN(CONTINUED)	SD(CONTINUED)
ACTG	.720	.449
ANTH	.284	.451
BA	.404	.491
BI	.511	.500
CH	.128	.334
EC	.680	.466
ENG	.349	.477
FHS	.299	.458
GEOL	.374	.484
HIST	.378	.485
J	.358	.479
MATH	.710	.454
MUS	.291	.454
PHIL	.268	.443
PHYS	.355	.479
PS	.346	.476
PSY	.318	.466
SOC	.339	.474
SPAN	.780	.414
WR	.601	.490

Some subjects such as Spanish, Accounting, and Math have relatively high continuation rates. These subjects contain more structured and specific sequences and people are most likely taking them with the intention of continuing with the rest of the sequence. For example, MATH 251 must be taken before you move onto 252. In the case of Math and Spanish, one year of mathematics is required for a Bachelor of Science and two years of foreign language are required for a Bachelor of Arts. This could also explain these high continuation rates because student likely need more than just one term in these subject areas to graduate. Others are quite low, such as Journalism, Sociology, and Psychology. This may be because students in these courses decide on another major and switch their direction of study.

The following table reports the number of observations, the mean, and standard deviation values for all the variables we will use in the regressions:

VARIABLE	OBS	MEAN	STD. DEV.	MIN	MAX
GED	489,989	.001	.038	0	1
HS NO TRANSFER HRS	489,989	.582	.493	0	1
HS WITH TRANSFER HRS	489,989	.173	.378	0	1
TRANSFER HRS, 1 TO 11	489,989	.006	.079	0	1
TRANSFER HRS, 12 TO 35	489,989	.027	.161	0	1
TRANSFER HRS, 36 TO 44	489,989	.028	.166	0	1
TRANSFER HRS, 45 TO 89	489,989	.104	.305	0	1
TRANSFER HRS, 90 TO 134	489,989	.070	.255	0	1
TRANSFER HRS, 135 OR MORE	489,989	.009	.093	0	1
LATINO	489,989	.066	.249	0	1
NATIVE AMERICAN	489,989	.008	.089	0	1
ASIAN	489,989	.054	.226	0	1
BLACK	489,989	.015	.123	0	1
PACIFIC ISLANDER	489,989	.006	.078	0	1
WHITE	489,989	.702	.458	0	1
BIRACIAL	489,989	.036	.186	0	1
NONRESIDENT ALIEN	489,989	.078	.269	0	1
UNKNOWN RACE	489,989	.035	.183	0	1
ACT	398,717	24.153	3.813	10	36
RESIDENT	489,989	.614	.487	0	1
INTERNATIONAL	489,989	.080	.271	0	1
STUDENT_MALE	489,989	.465	.499	0	1
PELL	489,989	.284	.451	0	1
CONTINUED	489,989	.444	.497	0	1
VALUE ADDED	206,700	-.040	.358	-4.022	2.875

INSTRUCTOR MALE	391,461	.628	.483	0	1
OTHER RANK	489,989	.021	.144	0	1
NO RANK	489,989	.032	.175	0	1
GE	489,989	.190	.392	0	1
UNKNOWN RANK	489,989	.244	.429	0	1
TTF	489,989	.241	.427	0	1
NTTF	489,989	.273	.445	0	1

First, we have student type, which consists of GED, graduated high school students without college credit, graduated high school students with college credit and transfer students from other colleges with varying levels of credits. We also have student race (we made a dummy variable for each race), which includes Latino, native American, Asian, black, pacific islander, white, non-resident alien unknown race and biracial. Other student demographic measurement we use is whether they are a resident or not and their gender, both of which we created dummy variables for. Finally, we know if they are an international student, and we know the ACT scores for any student that took the test. For students who took the SAT and not the ACT, we converted the SAT to the equivalent ACT score. We did this through the College Board Conversion formula. As for instructor characteristics, we use their gender, as well as their instructor level. We divided the instructor rank into tenured track faculty (TTF), non-tenured track faculty (NTTF), GE's, no rank, unknown rank, and all others (random, infrequent titles). We will use these to see how their rank and tenure-rank affect student learning.

The following are the 10 classes with the highest enrollment and the 10 most common pairs of classes in our data.

Course code	**MEAN(CONTINUED)**	**SD(CONTINUED)**
BA101	.388	.487
EC201	.722	.448
EC202	.660	.474
J201	.431	.495
MATH111	.787	.409
MATH243	.316	.465
PSY201	.304	.460
PSY202	.316	.465
WR121	.831	.375
WR122	.021	.144

Course pair	**Freq.**	**Percent**	**Cum.**
ACTG211/ACTG213	8,546	11.01	11.01
BA101/BA215	4,839	6.23	17.25
BI211/BI212	3,668	4.73	21.97
EC201/EC202	13,783	17.76	39.73
MATH111/MATH112	5,687	7.33	47.06
MATH111/MATH241	3,739	4.82	51.87
PHYS201/PHYS202	3,491	4.50	56.37
SPAN101/SPAN102	4,604	5.93	62.30
WR121/WR122	25,313	32.61	94.92
WR121/WR123	3,944	5.08	100.00
Total	77,614	100.00	

Methodology:

Step 1:

Our first set of regressions, in Table 3, were ran to determine the effect of different variables on our continued dummy variable. We ran this on our entire dataset (of 20 subjects), in order to get a true effect of certain variables on a student's likelihood to continue on in that subject. The first regression of this set was purely on subject, because we wanted to see the variance in continuation rates across subjects, as mentioned above. We essentially created a dummy variable for each subject to do this.

$$\text{Continuation} = \beta_0 + \beta_1\text{Subject}$$

We then added in professor characteristics (tenured/non-tenured, and gender), to see how they marginally affected continuation rate. (Regression 2) We used dummy variables for each rank, listed above.

$$\text{Continuation} = \beta_0 + \beta_1\text{Subject} + \beta_2\text{Professor Characteristics}$$

We then dropped the professor characteristics and replaced them with student characteristics to see their marginal effect. For these we have dummy variables for student type (i.e. their transfer credit hours, from high school, or another college), whether or not they received a Pell grant and their ACT score.

$$\text{Continuation} = \beta_0 + \beta_1\text{Subject} + \beta_2\text{Student Characteristics}$$

Regression 4 includes subject, professor characteristics, and student characteristics in one regression to see how they all interact with each other. Regression 5 is the same, but we omit instructor gender to correct for possible collinearity between rank and gender

$$\text{Continuation} = \beta_0 + \beta_1\text{Subject} + \beta_2\text{Professor Characteristics} + \beta_3\text{Student Characteristics}$$

Step 2:

For the next set of regressions, in Table 4, we specifically measure value added, and use a very similar 4 step process as before. However, we also add in the Ratio Continued as an independent variable in each of them to control for the selection effect. For these regressions, we left out any students that received a 4.0 or higher in both classes. We did this because students who max out the grade point scale in both classes would calculate as having 0 value added, which may not necessarily be representative of the truth. Therefore, our four regressions will respectively look like:

$$\text{Value Added} = \beta_0 + \beta_1\text{ Continuation Rate} + \beta_2\text{Subject}$$

$$\text{Value Added} = \beta_0 + \beta_1\text{ Continuation Rate} + \beta_2\text{Subject} + \beta_3\text{Student Characteristics}$$

$$\text{Value Added} = \beta_0 + \beta_1\text{ Continuation Rate} + \beta_2\text{Subject} + \beta_3\text{Professor Characteristics}$$

Value Added = $\beta_0 + \beta_1$ Continuation Rate+ β_2Subject + β_3Professor Characteristics + β_4Student Characteristics

Step: 3

Once we specified these models with all the variables, we used the full regression outputs to calculate predicted Continuation and value-added measures given the subject, professor characteristics and student characteristics. We can then compare the predicted values to the actual values of faculty.

Results and Discussion:

Standard errors in parentheses
*** $p<0.01$, significantly different from 0 at the 1% level
** $p<0.05$, significantly different from 0 at the 5% level
* $p<0.1$, significantly different from 0 at the 10% level

Model 1: Differences in Continuation by Subject

N = 489,989, R^2 = .09

ANTH	-0.460***
	(0.00568)
BA	-0.328***
	(0.00555)
BI	-0.221***
	(0.00580)
CH	-0.705***
	(0.00727)
EC	-0.0462***
	(0.00556)
ENG	-0.386***
	(0.00575)
FHS	-0.443***

	(0.00684)
GEOL	-0.360***
	(0.00594)
HIST	-0.355***
	(0.00559)
J	-0.376***
	(0.00593)
MATH	-0.0112**
	(0.00539)
MUS	-0.452***
	(0.00590)
PHIL	-0.480***
	(0.00608)
PHYS	-0.379***
	(0.00604)
PS	-0.389***
	(0.00586)
PSY	-0.420***
	(0.00557)
SOC	-0.397***
	(0.00574)
SPAN	0.0797***
	(0.00638)
WR	-0.131***

In our first model of step one (figure above), we see that every subject has a statistically significant variation in continuation rate when compared to Accounting (our omitted subject). This means that every subject in itself has an influence over whether students continue on or not. They are all negative coefficients except for Spanish, which means accounting has the second highest continuation rate across subjects. This remains constant throughout the rest of our models.

Model 2: Differences in Continuation by Instructor Characteristics

N = 391,461, R^2 = .0784

No Rank	0.0122***
	(0.00458)
TTF	0.00889***
	(0.00234)
NTTF	-0.00886***
	(0.00223)
other rank	0.00851
	(0.00544)
instructor male	0.00461**
	(0.00185)
unknown rank	

In our second model for measuring different controls in continued rate, we added in instructor characteristics. The most important factors to note here would be that tenured track faculty have a positive correlation with continuation, whereas non-tenured track faculty have a negative correlation. We also see that male instructors have a significant positive relationship with continuation rate. In addition, we see the R squared drop when we add in instructor characteristics.

Model 3: Differences in Continuation by Student Characteristics

N = 398,717, R^2 = .1016

hs no transfer hrs	0.0945***
	(0.0326)
hs with transfer hrs	0.0624*
	(0.0326)
transfer hrs, 1 to 11	0.0297

		(0.0342)
	transfer hrs, 12 to 35	-0.00494
		(0.0330)
	transfer hrs, 36 to 44	-0.0170
		(0.0332)
	transfer hrs, 45 to 89	-0.0267
		(0.0328)
	transfer hrs, 90 to 134	-0.0473
		(0.0332)
	transfer hrs, 135 or more	-0.110***
		(0.0392)
	latino	-0.00332
		(0.00585)
	native american	0.00129
		(0.0108)
	asian	0.00239
		(0.00597)
	black	0.0223***
		(0.00822)
	pacific islander	0.00707
		(0.0115)
	white	0.0103**
		(0.00494)
	biracial	0.00485
		(0.00646)
	nonresident alien	-0.00232
		(0.0266)
	pell	0.0150***
		(0.00201)
	act	-0.00109***
		(0.000234)
	student male	0.00828***
		(0.00174)
	international	0.0304
		(0.0262)
	resident	-0.00249
		(0.00187)

Then we replaced instructor characteristics with student characteristics. Arguably the most important finding is our R squared, which increases from about .07 to about .1, indicating that the student characteristics explain the differences in our continued dummy variable much more than instructor characteristics. We find that only high school transfers with no credit hours, and transfers from other universities with over 135 credits are significant. The high school transfers have a positive correlation, and the university transfers a negative correlation. This negative correlation may be explained by the fact that they are so close to graduation that they only needed one more class in a given subject to meet their degree requirements. The only significant ethnicities were black and white students, both of which had positive effects on continuation (in relation to unknown ethnicity). Pell Grant recipients, as well as male students also have a significant and positive correlation. Pell Grant recipient effect was relatively large, whereas gender effect was small. Finally, we have ACT scores, which actually have a negative correlation with continuation rate. This negative correlation is pretty hard to explain. However, we have nearly 500,000 observations, so even the smallest effects will be captured. So even though it is statistically significant, the effect is extremely close to zero, and economically insignificant

Finally, we combined the student and instructor characteristics, but omitted instructors with no rank in one, and omitted male instructors in another, due to collinearity issues.

When we omit the unknown instructors (Model 4 of Table 1 in the appendix), we find the same significance with high school transfers with no credit hours and university transfers with over 135 credit hours as mentioned above. However, the only other significant student controls are black students and Pell Grant recipients (both positively correlated). We also see again that tenured track faculty have positive correlation, whereas non-tenured track has a negative effect on continuation. Finally, male is again significant with a positive correlation, and our "other rank" category is also. Here we get an R squared of .084, with 320,654 observations.

When we omit our male instructor dummy variable (Model 5 of Table 1 in the appendix), our student effects stay the same, except both male students and white students are both now significant and positively correlated with continuation rate. ACT is again significant now in this model, and again have a negative effect. The largest differences, as expected, are seen in instructor factors. Non-ranked faculty have a very significant negative correlation with continuation rate, and so do tenured and non-tenured now. In addition, the marginal effect is much larger than it was in our 4th model. We then ran some tests to measure how significant certain subjects were against each other, as opposed to accounting, as well as how significant the difference was between tenured track and non-tenured track faculty. As seen below from our 5th and final model, all mentioned variables remain significant. However, our R squared jumps to .1026, indicating this is a better fit than model 4, and has 389,717 observations.

- TTF - NTTF = 0, chi2(1) = 108.47, Prob > chi2 = 0.0000

- ENG - MATH =0, chi2(1) = 5977.02, Prob > chi2 = 0.0000

- EC - PSY= 0, chi2(1) = 6401.74, Prob > chi2 = 0.0000

Looking at our 5 models, we see a strong, consistent effect of subject variance. Then when we added in instructor characteristics, we see a significant, but small effect of TTF (positive) and NTTF (negative), which is consistent with our expectations.

When replaced with student characteristics, we see black and white students both have a large and significant effect on continuation, which remains constant throughout our combined model as well. Pell recipients also has an effect that remains unchanged in our 4th and 5th model as well. Finally, we have our gender effect, which has a small yet significant effect.

Then in our combined model without our faculty gender variable, all of our previous findings remain constant except for TTF. This now has a much larger effect, and in fact becomes a negative effect. Then, when we keep it, white students become insignificant, TTF becomes positive again, and male instructors becomes significant with a positive effect again.

Step 2: Value Added Models

Model 1: Value Added by Ratio Continued and Subject

N = 146,376, R^2 = .0029

Ratio Continued	-0.0126*
	(0.00720)
ANTH	-0.000387
	(0.00679)
BA	0.00553
	(0.00617)
BI	-0.0345***
	(0.00595)
CH	-0.110***
	(0.0108)
EC	0.0135***
	(0.00508)
ENG	0.0116*
	(0.00650)
FHS	0.0185
	(0.0116)
GEOL	0.00890
	(0.00681)
HIST	0.0147**
	(0.00607)
J	-0.00896
	(0.00707)
MATH	-0.0349***
	(0.00498)
MUS	0.0130
	(0.00934)
PHIL	-7.47e-05
	(0.00772)
PHYS	-0.0205***
	(0.00706)
PS	-0.00438
	(0.00703)
PSY	-0.00884

	(0.00647)
SOC	-0.00843
	(0.00678)
SPAN	0.00300
	(0.00569)
WR	0.00714
	(0.00493)

In the first model of our value added (step 2) regressions (Table above), only some of the subjects are statistically significant. Biology, Chemistry, Economics, Mathematics and Physics have the highest level of significance when compared to Accounting. All of these have a negative coefficient with respect to value added when compared to Accounting with the exception of Economics. These negative correlations are probably indicative of the difficulty of these subjects. The academic rigor of these math-heavy courses likely makes it difficult for most students to have upward mobility in comparison with their peers regardless of the teaching quality. English and History are also statistically significant with positive coefficients.

The coefficient for Ratio Continued in this model is negative and statistically significant. This can be interpreted as meaning that the more students that continue on from a given class the lower the average value added will be for the future class. This is in line with the prediction that a selection effect could be affecting the value-added measure.

Model 2: Value Added by Student Characteristics

N = 122,261, R^2 = .0053

hs no transfer hrs	0.0423
	(0.0477)
hs with transfer hrs	0.0471
	(0.0477)
transfer hrs, 1 to 11	0.0413
	(0.0495)
transfer hrs, 12 to 35	0.0439
	(0.0482)
transfer hrs, 36 to 44	0.0563
	(0.0483)
transfer hrs, 45 to 89	0.0717
	(0.0480)
transfer hrs, 90 to 134	0.0534
	(0.0484)
transfer hrs, 135 or more	0.00281
	(0.0581)
latino	0.00467
	(0.00707)
native american	-0.0269**
	(0.0127)
asian	0.00114
	(0.00729)
black	-0.0140
	(0.00952)
pacific islander	0.00508
	(0.0136)
white	-0.00267
	(0.00608)
biracial	0.00247
	(0.00787)
nonresident alien	0.00697
	(0.0346)
act	-0.00335***
	(0.000299)
resident	-0.0112***
	(0.00223)
international	-0.00671
	(0.0341)
pell	-0.00391

	(0.00240)
student male	-0.00746***
	(0.00209)

In the second model, where we added in student characteristics, all of these subjects maintained the direction of their coefficients and their level of significance with the exception of Physics and Music. Physics is no longer statistically significant while Music has gained statistical significance. The Ratio Continued coefficient is now positive but is no longer significant. This suggests that the student characteristics account for some of the selection affect and absorb some of the effect of the subject variance. The coefficients for student types are all positive when compared to the omitted dummy variable, GED, but none are statistically significant. The only race that is statistically significant is Native American and has a negative coefficient in comparison to the omitted dummy, unknown race. The coefficients for ACT, resident dummy and student male dummy are all statistically significant and are negatively correlated.

Model 3: Value Added by Instructor Characteristics

N = 108,869, R^2 = .0026

No Rank	-0.000180
	(0.00594)
TTF	0.00281
	(0.00316)
NTTF	-0.000578
	(0.00299)
other rank	-0.00942
	(0.00696)
instructor male	-0.00413*
	(0.00246)
unknown rank	

In the third model, which contains instructor characteristics but not the student characteristics, some of the coefficients for the subjects maintain their direction and significance from the first model: Biology, Chemistry, Math and Physics. Economics, English and History are no longer statistically significant. Anthropology, Journalism, Psychology and Sociology are now significant with negative correlation. None of the instructor characteristics are statistically significant with the exception of the instructor male dummy variable which has a negative correlation. Like the continuation models, there is collinearity between the instructor rank dummy variables and instructor gender due to missing data. Therefore, both GEs and unknown ranked instructors have been omitted. Here we see again, that student characteristics are more explanatory then instructor characteristics.

In the fourth and fifth models (Models 4 and 5 of Table 2 in the appendix) we ran regressions with the subjects and both the student and the instructor characteristics. The difference between the two is the fourth omits the dummy variable for unknown rank and the fifth omits the dummy variable for instructor male to deal with the collinearity issue. The Ratio Continued variable is no longer statistically significant in these models. The coefficients for ACT, resident dummy and student male dummy are still statistically significant and are negatively correlated in both of these models. The coefficients for the dummy variables for TTF (positive) and instructor male(negative) are significant in the fourth model. TTF is not significant in the fifth model and instructor male is not included. We get 90,871 observations, and an R squared of .0053 for our 4th model.

For the fifth model we ran a few tests to test the difference between a few variables. First, we tested for significantly different coefficients between tenured track faculty and non-tenured track faculty. We have 122,261, and an R squared of .0057 for this model. This shows there is little difference when we omit instructor gender as opposed to one of their ranks.

- TTF - NTTF = 0, F(1,122214) = 0.14, Prob > F =0.7049

The difference between these two instructor types is not statistically significant. Next we tested the difference between some sets of subjects:

- CH + MUS = 0, F(1,122214) = 89.08, Prob > F =0.0000

- CH + ENG = 0, F(1,122214) = 108.82, Prob > F = 0.0000

- MATH + MUS = 0, F(1,122214) = 45.69, Prob > F = 0.0000

ENG - MATH = 0, F(1,122214) = 91.32, Prob > F = 0.0000

- CH + EC = 0, F(1,122214) = 101.21, Prob > F = 0.0000

All of these tests show that these pairs are significantly different from each other.

Biology, Chemistry and Math are statistically significant and have negative coefficients in all the value-added regressions. ACT score, the resident dummy variable and the student male dummy variable maintain their statistical significance throughout all five models and have negative coefficients. Again, the negative coefficient for ACT score is peculiar and contrary to what one would expect. As with the continuation model, this is likely explained by the number of observations: the negative coefficient is statistically significant but is rather small. With the instructor rank dummy variables, almost all are not significantly different from the omitted groups. These coefficients are inconsistent across models and even switch signs. This gives the general impression that individual ranks are not very explanatory when it comes to measuring value added.

Conclusion:

In their paper, Ancell and Wu used Journalism and Business data to call into question the validity of SETs as accurate measures of teaching quality. This is in line with much of the academic research on SETs. The general conclusions about SETs is that they are biased based on gender, race, physical appearance and grading leniency. This is why constructed a more objective measure of teaching quality.

The results from the value-added shows that there aren't any consistent conclusions that can be made about the teaching quality of instructors based on their rank category. This is evident that these positions should not lead to immediate decisions about an instructor based on their rank and instructor type. Instead, their merit as teachers should base off a measure of teaching quality such as the value-added measure. Therefore, using this measure in this manner becomes very useful for the University and the individual departments.

This measure can be used to narrow down which instructors promote continuation with a subject. This is especially valuable to departments with low and decreasing enrollment. Using the continuation measures to combat discouragement within a subject could provide quantitative measures to inform policy about promoting these majors that suffer from decreasing interest.

There are some downsides to these measures that we created. The way we calculated the value-added measure didn't allow us to account for cross-teaching effects. This would occur when a student took multiple classes within a subject in the same term. For these instances, it would be impossible to determine which instructor contributed to the student's value in future courses. Also, using these measures to determine teacher merit and promotions would lead to unintended behaviors. It is likely that a teacher would respond to the implementation of these measures by changing their strategy for instruction. They would have incentives to pass as many students as possible to increase their continuation rates, and to teach to the next class to increase their value-added score. This could lead to students receiving inadequate education from instructors that are just trying to pass them all and only provide them with the necessary information for the next course.

The models we created have another use that could be very useful for policy-making. We used our models to create predicted values for continuation rates and value-added controlling for the other characteristics of the instructors, students and classes that we used. This means that we were able to subtract a predicted measure from an instructor's actual measure. This creates a measure of teaching success controlling for the effects of the external factors that affect an instructor's students. This leaves behind the immeasurable qualities of the instructor: their level of teaching ability. A summary of these two variables (actual continuation minus predicted continuation, actual value added minus predicted value added) are included in the Appendix labeled Table 3. A list of these new measures for Economics professors in our data are included in Table 4 in the Appendix. This table shows that there is a large variance in both measures across the Economics instructors. There are few instructors that have positive values in the differences between the actual measure and the predicted measure for both models. These two facts are evidence that these measures are able to reflect differences in teaching quality across faculty members.

The failure of SETs as measures of teaching begs for a new and more accurate way of determining instructor's abilities as teachers. We created a more objective and accurate model to measure an instructor teaching ability. For future research we suggest comparing SET scores to the value-added measures. We also suggest doing this comparison to try to determine the biases that exist in SETs by using these models to control for teaching quality.

Appendix:

Table 1

Continuation rate model

		(1)	(2)	(3)	(4)	(5)
	VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5
	ANTH	-0.460***	-0.441***	-0.487***	-0.463***	-0.480***
		(0.00568)	(0.00632)	(0.00665)	(0.00739)	(0.00668)
	BA	-0.328***	-0.313***	-0.386***	-0.367***	-0.375***
		(0.00555)	(0.00613)	(0.00655)	(0.00725)	(0.00661)
	BI	-0.221***	-0.201***	-0.245***	-0.225***	-0.238***
		(0.00580)	(0.00640)	(0.00677)	(0.00748)	(0.00680)
	CH	-0.705***	-0.688***	-0.762***	-0.744***	-0.760***
		(0.00727)	(0.00779)	(0.00847)	(0.00908)	(0.00852)
	EC	-0.0462***	-0.0420***	-0.0631***	-0.0655***	-0.0717***
		(0.00556)	(0.00648)	(0.00660)	(0.00764)	(0.00668)
	ENG	-0.386***	-0.377***	-0.405***	-0.394***	-0.409***
		(0.00575)	(0.00659)	(0.00672)	(0.00768)	(0.00676)
	FHS	-0.443***	-0.420***	-0.461***	-0.433***	-0.458***
		(0.00684)	(0.00741)	(0.00788)	(0.00857)	(0.00791)
	GEOL	-0.360***	-0.351***	-0.381***	-0.370***	-0.378***
		(0.00594)	(0.00666)	(0.00689)	(0.00775)	(0.00694)
	HIST	-0.355***	-0.345***	-0.369***	-0.355***	-0.369***
		(0.00559)	(0.00628)	(0.00656)	(0.00736)	(0.00666)
	J	-0.376***	-0.366***	-0.392***	-0.376***	-0.389***
		(0.00593)	(0.00656)	(0.00689)	(0.00763)	(0.00696)
	MATH	-0.0112**	-0.0317***	-0.0183***	-0.0409***	-0.0284***
		(0.00539)	(0.00635)	(0.00643)	(0.00759)	(0.00646)
	MUS	-0.452***	-0.450***	-0.488***	-0.486***	-0.483***
		(0.00590)	(0.00656)	(0.00694)	(0.00771)	(0.00696)
	PHIL	-0.480***	-0.456***	-0.507***	-0.483***	-0.511***
		(0.00608)	(0.00681)	(0.00700)	(0.00784)	(0.00708)
	PHYS	-0.379***	-0.360***	-0.406***	-0.386***	-0.396***

	(0.00604)	(0.00657)	(0.00709)	(0.00774)	(0.00713)
PS	-0.389***	-0.373***	-0.408***	-0.391***	-0.410***
	(0.00586)	(0.00662)	(0.00685)	(0.00774)	(0.00688)
PSY	-0.420***	-0.407***	-0.465***	-0.445***	-0.464***
	(0.00557)	(0.00632)	(0.00655)	(0.00738)	(0.00664)
SOC	-0.397***	-0.380***	-0.427***	-0.409***	-0.425***
	(0.00574)	(0.00641)	(0.00671)	(0.00747)	(0.00679)
SPAN	0.0797***	0.0904***	0.0848***	0.0949***	0.0855***
	(0.00638)	(0.00747)	(0.00749)	(0.00873)	(0.00751)
WR	-0.131***	-0.162***	-0.147***	-0.183***	-0.158***
	(0.00518)	(0.00622)	(0.00612)	(0.00729)	(0.00617)
No Rank		0.0122***		0.00305	-0.0485***
		(0.00458)		(0.00516)	(0.00526)
TTF		0.00889***		0.0110***	-0.0331***
		(0.00234)		(0.00258)	(0.00295)
NTTF		-0.00886***		-0.00566**	-0.0599***
		(0.00223)		(0.00250)	(0.00268)
other rank		0.00851		0.0159***	-0.0342***
		(0.00544)		(0.00610)	(0.00626)
instructor male		0.00461**		0.00669***	
		(0.00185)		(0.00206)	
unknown rank					-0.0467***
					(0.00296)
hs no transfer hrs			0.0945***	0.102***	0.0960***
			(0.0326)	(0.0366)	(0.0326)
hs with transfer hrs			0.0624*	0.0795**	0.0635*
			(0.0326)	(0.0366)	(0.0327)
transfer hrs, 1 to 11			0.0297	0.0529	0.0310
			(0.0342)	(0.0383)	(0.0343)
transfer hrs, 12 to 35			-0.00494	0.0364	-0.00384
			(0.0330)	(0.0370)	(0.0331)
transfer hrs, 36 to 44			-0.0170	0.0277	-0.0150
			(0.0332)	(0.0371)	(0.0332)
transfer hrs, 45 to 89			-0.0267	0.00908	-0.0254

				(0.0328)	(0.0368)	(0.0329)
	transfer hrs, 90 to 134			-0.0473	-0.00657	-0.0471
				(0.0332)	(0.0372)	(0.0332)
	transfer hrs, 135 or more			-0.110***	-0.0875**	-0.109***
				(0.0392)	(0.0438)	(0.0393)
	latino			-0.00332	-0.00220	0.000270
				(0.00585)	(0.00645)	(0.00586)
	native american			0.00129	0.00358	0.00349
				(0.0108)	(0.0118)	(0.0108)
	asian			0.00239	0.000894	0.00462
				(0.00597)	(0.00659)	(0.00597)
	black			0.0223***	0.0203**	0.0247***
				(0.00822)	(0.00901)	(0.00823)
	pacific islander			0.00707	0.00255	0.00940
				(0.0115)	(0.0125)	(0.0115)
	white			0.0103**	0.00744	0.0125**
				(0.00494)	(0.00549)	(0.00495)
	biracial			0.00485	0.00663	0.00835
				(0.00646)	(0.00707)	(0.00646)
	nonresident alien			-0.00232	0.00191	0.00255
				(0.0266)	(0.0305)	(0.0267)
	pell			0.0150***	0.0151***	0.0151***
				(0.00201)	(0.00219)	(0.00201)
	act			-0.00109***	0.000168	-0.00107***
				(0.000234)	(0.000254)	(0.000234)
	student male			0.00828***	0.00182	0.00810***
				(0.00174)	(0.00189)	(0.00174)
	international			0.0304	0.0270	0.0282
				(0.0262)	(0.0302)	(0.0263)
	resident			-0.00249	-0.00219	-0.00343*
				(0.00187)	(0.00203)	(0.00187)
	Observations	489,989	391,461	398,717	320,654	398,717

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 2

Value added model

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5
ratio continued	-0.0126*	0.00305	-0.0205**	-0.00859	0.00322
	(0.00720)	(0.00814)	(0.00816)	(0.00911)	(0.00817)
ANTH	-0.000387	0.00323	-0.0126*	-0.00825	0.00283
	(0.00679)	(0.00733)	(0.00746)	(0.00802)	(0.00737)
BA	0.00553	0.00402	-0.000739	0.000684	0.00430
	(0.00617)	(0.00685)	(0.00675)	(0.00747)	(0.00691)
BI	-0.0345***	-0.0362***	-0.0422***	-0.0425***	-0.0367***
	(0.00595)	(0.00641)	(0.00656)	(0.00708)	(0.00650)
CH	-0.110***	-0.104***	-0.120***	-0.111***	-0.104***
	(0.0108)	(0.0119)	(0.0114)	(0.0126)	(0.0120)
EC	0.0135***	0.0124**	-0.00218	-0.00372	0.0116**
	(0.00508)	(0.00554)	(0.00615)	(0.00669)	(0.00569)
ENG	0.0116*	0.0175**	0.00395	0.0123	0.0171**
	(0.00650)	(0.00700)	(0.00757)	(0.00812)	(0.00707)
FHS	0.0185	0.0183	0.0117	0.0143	0.0179
	(0.0116)	(0.0125)	(0.0121)	(0.0130)	(0.0125)
GEOL	0.00890	0.00820	0.00713	0.00818	0.00841
	(0.00681)	(0.00727)	(0.00770)	(0.00825)	(0.00734)
HIST	0.0147**	0.0142**	0.00631	0.00739	0.0132**
	(0.00607)	(0.00654)	(0.00682)	(0.00734)	(0.00669)
J	-0.00896	-0.00611	-0.0176**	-0.0149*	-0.00717
	(0.00707)	(0.00758)	(0.00772)	(0.00828)	(0.00765)
MATH	-0.0349***	-0.0427***	-0.0268***	-0.0361***	-0.0426***
	(0.00498)	(0.00544)	(0.00610)	(0.00672)	(0.00548)
MUS	0.0130	0.0220**	0.00225	0.0125	0.0216**
	(0.00934)	(0.00994)	(0.00975)	(0.0104)	(0.00998)
PHIL	-7.47e-05	0.0116	-0.0138	-0.00127	0.0109
	(0.00772)	(0.00817)	(0.00859)	(0.00909)	(0.00833)
PHYS	-0.0205***	-0.00801	-0.0294***	-0.0147*	-0.00836
	(0.00706)	(0.00771)	(0.00758)	(0.00826)	(0.00775)

PS	-0.00438	0.00410	-0.00614	0.00243	0.00372
	(0.00703)	(0.00754)	(0.00798)	(0.00854)	(0.00759)
PSY	-0.00884	-0.00228	-0.0205***	-0.0136*	-0.00318
	(0.00647)	(0.00712)	(0.00735)	(0.00798)	(0.00726)
SOC	-0.00843	-0.00808	-0.0194**	-0.0188**	-0.00893
	(0.00678)	(0.00734)	(0.00752)	(0.00809)	(0.00748)
SPAN	0.00300	0.000493	-0.00336	-0.00444	0.000368
	(0.00569)	(0.00612)	(0.00687)	(0.00737)	(0.00616)
WR	0.00714	0.00752	0.000483	0.00441	0.00762
	(0.00493)	(0.00533)	(0.00628)	(0.00677)	(0.00542)
hs no transfer hrs		0.0423		0.0373	0.0427
		(0.0477)		(0.0523)	(0.0477)
hs with transfer hrs		0.0471		0.0421	0.0474
		(0.0477)		(0.0523)	(0.0477)
transfer hrs, 1 to 11		0.0413		0.0354	0.0416
		(0.0495)		(0.0543)	(0.0495)
transfer hrs, 12 to 35		0.0439		0.0339	0.0443
		(0.0482)		(0.0528)	(0.0482)
transfer hrs, 36 to 44		0.0563		0.0424	0.0566
		(0.0483)		(0.0530)	(0.0483)
transfer hrs, 45 to 89		0.0717		0.0650	0.0720
		(0.0480)		(0.0526)	(0.0480)
transfer hrs, 90 to 134		0.0534		0.0424	0.0538
		(0.0484)		(0.0531)	(0.0484)
transfer hrs, 135 or more		0.00281		0.0286	0.00310
		(0.0581)		(0.0652)	(0.0581)
latino		0.00467		0.00451	0.00472
		(0.00707)		(0.00813)	(0.00708)
native american		-0.0269**		-0.0127	-0.0269**
		(0.0127)		(0.0144)	(0.0127)
asian		0.00114		-0.000248	0.00115
		(0.00729)		(0.00842)	(0.00729)
black		-0.0140		-0.0107	-0.0140
		(0.00952)		(0.0109)	(0.00952)
pacific islander		0.00508		0.00475	0.00504

		(0.0136)		(0.0155)	(0.0136)
white		-0.00267		-0.00167	-0.00262
		(0.00608)		(0.00704)	(0.00608)
biracial		0.00247		0.00104	0.00252
		(0.00787)		(0.00901)	(0.00788)
nonresident alien		0.00697		0.0229	0.00683
		(0.0346)		(0.0425)	(0.0346)
act		-0.00335***		-0.00339***	-0.00336***
		(0.000299)		(0.000339)	(0.000300)
resident		-0.0112***		-0.00859***	-0.0112***
		(0.00223)		(0.00254)	(0.00223)
international		-0.00671		-0.0205	-0.00654
		(0.0341)		(0.0420)	(0.0341)
pell		-0.00391		-0.00500*	-0.00391
		(0.00240)		(0.00272)	(0.00240)
student male		-0.00746***		-0.00931***	-0.00748***
		(0.00209)		(0.00238)	(0.00209)
No Rank			-0.000180	0.00524	0.00757
			(0.00594)	(0.00619)	(0.00604)
TTF			0.00281	0.00581*	0.00288
			(0.00316)	(0.00334)	(0.00345)
NTTF			-0.000578	-0.000439	0.00166
			(0.00299)	(0.00319)	(0.00298)
other rank			-0.00942	-0.000182	0.00135
			(0.00696)	(0.00736)	(0.00730)
instructor male			-0.00413*	-0.00499*	
			(0.00246)	(0.00260)	
unknown rank					0.000472
					(0.00348)
Constant	-0.0279***	0.00631	-0.0130	0.0263	0.00482
	(0.00688)	(0.0493)	(0.00818)	(0.0543)	(0.0493)
Observations	146,376	122,261	108,869	90,871	122,261
R-squared	0.003	0.006	0.003	0.005	0.006

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3

Summary of differences between actual and predicted for continuation rates and value added

Variable	Obs	Mean	Std. Dev.	Min	Max
diff_cont	1,402	-.0176306	.2089965	-.8118139	.9230908

Variable	Obs	Mean	Std. Dev.	Min	Max
diff_val	1,286	.0135275	.1332563	-.9110349	1.188647

Table 4

Summary of Economics professors

RANK	CONTINUED	VALUE ADDED	SALARY	DIFF IN CONT.	DIFF IN VALUE ADDED
	0.4375	-0.0438888	96733	-0.1838405	-0.0177769
	0.3157895	-0.1672149	4020	-0.3522151	-0.1379306
	0.4285714	-0.21729	64998	-0.2189923	-0.1736213
	0.5208333	-0.0671523	1500	-0.1082185	-0.0290486
	0.6363636	-0.0176033	51453	-0.0574585	0.0131062
	0.6739131	0.0097397	202433	0.0615997	0.039465
	0.56	-0.0168895	108054	-0.0384005	0.0100331

	0.25	-0.0039029	60	-0.3646726	0.0379664
	0.615	0.0064014	4000	-0.0403232	0.0349791
	0.2222222	-0.0191061	153425	-0.4465038	0.0155932
	0.52	0.0821247	2000	-0.0986402	0.1140841
	0.698708	-0.0271467	38097	0.0155067	0.0009607
	0.5	0.0355269		-0.1297206	0.0629923
Assistant Professor	0.2	0.1067446	112200	-0.4385507	0.1311132
Assistant Professor	0.6591761	-0.0188668	112200	-0.0271379	0.0092718
Assistant Professor	0.3223141	-0.0286363	96733	-0.316566	0.0002566
Assistant Professor	0.3333333	-0.0548381	84159	-0.3558699	-0.0310429
Assistant Professor	0.7627315	-0.0432208	88297	0.0706115	-0.0142848
Assistant Professor	0.7659575	-0.0277886		0.0170344	0.015218
Assistant Professor	0.2666667	-0.0779444	105195	-0.3812	-0.0398322
Assistant Professor	0.770389	-0.047645	96733	0.0734785	-0.0200321
Assistant Professor	1	0.1100494	103000		
Assistant Professor	1	0.1744383	61950	0.420409	0.2237118
Assistant Professor	0		96733	-0.5999429	
Assistant Professor	0.632312	0.0165184	108246	-0.052268	0.0460329
Associate Professor	0.7514231	-0.0297738	102267	0.0556803	-0.0014468
Associate Professor	0.7668594	-0.0665102	97449	0.0755987	-0.0380655
Associate Professor	0.4210526	-0.051458	90000	-0.1699534	-0.0364115
Associate Professor	0.8333333	-0.0701527	70000	0.1194127	-0.0350697
Associate Professor	0.5227273	0.0481398	85942	-0.1318458	0.0792491
Associate Professor	0.6895834	0.002737	101835	-0.0045157	0.0304562
Instructor	0.7142857	-0.0408502	60000	0.0397163	-0.0107159
Instructor	0.7617021	0.0735199	82620	0.0859582	0.1003187
Instructor	0.6268657	0.079275	55080	-0.0241813	0.1060021
Instructor	1	-0.1135796	68850		
Instructor	0.4509804	-0.0028606	65000	-0.2176125	0.0264061
Instructor	0.6080178	-0.0218137	52000	-0.0667957	0.0083643
Instructor	0.7802469	0.0207485	68850	0.1080012	0.0515832
Instructor	0.7241379	0.1081209	55080	0.0588432	0.1385789
Instructor	0.5	-0.0568111	70000	-0.0813788	-0.0678926
Instructor	1	0.0094937	81600	0.4272938	0.0248356
Instructor	0.6	-0.0007497	60003	-0.0368635	0.0306503
Instructor	0.7340659	0.0439966	21600	0.0564518	0.0741156
Instructor	0.25	-0.301035	55080	-0.3194764	-0.271729

Instructor	0.5172414	-0.3372605	55080	-0.1010814	-0.3168491
No Rank	0.6531414	-0.0407347	136146	-0.0224648	-0.0181047
No Rank	1	-0.0560558	67500		
No Rank	0.5	0.0772498	42545	-0.1133106	0.1006935
No Rank	0.3333333	0.4233205	150965	-0.316424	0.4767636
Postdoctoral Research Assoc	0.5555556	-0.1177046	33974	-0.0729491	-0.0903172
Professor	0.6081921	-0.0276297	119358	-0.0815923	0.0008528
Professor	0.3584416	-0.0571311	105962	0.0322066	-0.0171354
Professor	0.5882353	-0.0762107	73500	-0.070174	-0.0462233
Professor	0.3333333		186662	-0.365248	
Professor	0.7547893	0.0046228	118165	0.0565429	0.0365802
Professor	0.5352113	-0.1035715	117020	-0.0953919	-0.0779074
Professor	0.3947369	-0.0708214	111327	-0.2479266	-0.0405881
Professor	0.6213592	-0.1151663	120288	-0.0753808	-0.0861712
Professor of Practice	0.6194388	-0.0835162	94969	-0.0448362	-0.0546587
Research Assistant	0.7416546	-0.0265	45934	0.0478721	0.0058973
Research Assistant	0.5384616	-0.296492	48895	-0.0867268	-0.2852046
Research Associate	0.6153846	-0.0799895	70000	-0.0314464	-0.0527641
Senior Instructor	0.6666667	0.1217106	44890		
Senior Instructor-LEGACY	0.5588235	-0.1120037	62369	0.1894611	-0.0831256