Institutional Research and Effectiveness

Working Paper Series

Measuring Teaching Effectiveness using SETs

LLOYD A. BLANCHARD

Department of Public Policy, University of Connecticut

YIZHI ZHU

Department of Economics, University of Connecticut

June 2019
Working Paper No. 2

For correspondence contact:
Yizhi Zhu, yizhi.zhu@uconn.edu, (860) 486-4240

Prepared for presentation at the
61st Annual Western Social Science Association Conference
San Diego, CA April 24–27, 2019

This is a preliminary draft. Comments are very welcome.

Abstract

Student Evaluation of Teaching (SET) surveys are used commonly in colleges and universities, from community colleges to Ivy League in the United States to institutions of higher education in other countries, such as Canada, Australia, England and China (Marsh and Roche, 1997). Typically, the SET survey includes around 20 questions on multiple dimensions of teaching and the course. The purpose of this paper is to examine the validity of SETs in capturing teaching effectiveness by applying two different empirical methods to SET data from a large public research university. Our hypothesis is that SET feedback captures teaching effectiveness in the form of learning. In other words, the more students learn, the higher are the instructor and course SET ratings. Our first method analyzes individual student SETs as the unit of analysis, relating SET ratings to subjective measures of learning (i.e., what the student say they learned). We find support for our hypothesis using this approach. Our second method uses a multi-level regression model designed by Weinberg et al (2009) that places course sections (subject by subject) as the unit of analysis, relating SET ratings to objective measures of learning based on the link between performance in prerequisite courses (e.g., ECON 101, PSY 101) and subsequent courses (e.g., ECON 201, PSY 201). We find little evidence in support of our hypothesis using this second approach. We conclude that universities should give proper weight to SET results, especially if they affect personnel decisions. However, subjective measures of learning should not be discounted out of hand, as students who say they are learning may well be doing so.

Measuring Teaching Effectiveness using SETs

1. Introduction

Student Evaluation of Teaching (SET) surveys are used commonly in colleges and universities, from community colleges to Ivy League in the United States to institutions of higher education in other countries, such as Canada, Australia, England and China (Marsh and Roche, 1997). At the end of each semester, students across the world are asked (and sometimes incentivized or compelled) to rate their courses and instructors, and to the extent that SET evaluations are used for accountability by college administrators and impact on instructors' employment, one should be convinced that such evaluations measure teaching effectiveness accurately, consistently, and fairly. Many university faculty are not convinced they are.

Typically, the SET survey includes around 20 questions on multiple dimensions of teaching and the course. Some institutions adopt SET instruments such as the Students' Evaluations of Educational Quality (SEEQ) that are validated scientifically and comprehensively used by other institutions, while others choose to develop and validate their own SET instruments. Marsh (2007) claims that SET surveys mainly serve five purposes. First, they universally provide diagnostic feedback to faculty, department heads, deans, and other academic administrators charged with maintaining teaching quality. Second, they are tools for the measurement of teaching effectiveness for personnel decisions. This second purpose is highly controversial, as faculty are loath to have their employment, promotion, and/or tenure prospects influenced by the judgment (or whim) of students. SETs are used for the other three purposes on a more sporadic basis. Third, they provide students information for selecting courses and instructors. Fourth, they contribute to accreditation and quality assurance efforts designed to monitor teaching quality and learning. Fifth, they provide the results for research on teaching effectiveness.

It seems likely that one's position on the use of SETs depends strongly on how high the stakes are. Indeed, SETs are useful tools to help improve teaching and our understanding of teaching effectiveness, but they should face heightened scrutiny (if used at all) when employment decisions are involved, and this scrutiny should focus on the accuracy, consistency, and fairness with which the SET survey instrument measures teaching effectiveness.

Accuracy relates to the extent to which the SET instrument measures teaching effectiveness, which is a complex concept presumably comprised of a number of dimensions. For example, the SEEQ instrument is designed to capture the following nine dimensions of teaching/course quality:

1. Learning and value
2. Instructor enthusiasm
3. Organization and clarity
4. Student interaction
5. Individual rapport
6. Breadth of coverage
7. Examinations and grading
8. Assignments and readings
9. Workload and difficulty

Consistency refers to whether the SET instrument is measuring the same thing for one instructor or course as another. That is, students who complete the surveys would need to apply the questions to their experience in a similar fashion. Finally, fairness refers to whether the results of the SET are correlated with race/ethnicity and/or gender. To the extent that SET evaluations treat instructors differently based on these characteristics, and such results are used in high stakes employment decisions, at best, this leaves serious employment decisions in the hands of young adults whose SET responses are sometimes serious, sometimes whimsical, and sometimes cruel; and at worst, it violates equal employment law.

The purpose of this paper is to examine the validity of SET by applying two different empirical methods to SET data from a large public research university. We use different methods to address some of the methodological issues inherent in student evaluation data, which we summarize next in our literature review.

2. Literature Review

In his meta-analysis and literature review, Clayson (2009) began by claiming "few issues within academics have been as well researched, documented, and long lasting as the debate about the student evaluation of teaching." A number of comprehensive literature reviews exist (Benton and Cashin, 2012; Abrami et al, 2007; Marsh, 2007; Cohen, 1981), so here we will focus mainly on the key issues to be considered. The key question is: "are the evaluations that students make of courses and instructors related to student learning?" There is no general consensus in the literature on this question of the validity of SETs in measuring teaching effectiveness. Clayson's (2009) meta-analysis found that a small average relationship exists between learning and SETs, but he concludes that the association is "situational and not applicable to all teachers, academic disciplines, or levels of instruction. It is concluded that the more objectively learning is measured, the less likely it is to be related to the evaluations."

Clayson (2009) highlighted three types of problems that confound research examining the relationship between learning and SETs. The first set of problems is situational, mainly in terms of the differences among the disciplines themselves. Most research on SETs emanates from educational researchers, who tend to have a more positive attitude toward the use of such instruments. Researchers in other disciplines, who often consider education research less than rigorous, are more skeptic. Different approaches to SET research are related to different disciplinary perspectives. Moreover, SET instruments may capture different phenomena in different disciplines, as not all teaching is alike. Pedagogy in physical sciences is certainly different than in humanities, and SETs are expected to capture learning across wildly different contexts.

The second set of problems is methodological, particularly in terms of how learning is measured. Most empirical research in this area seeks to find a statistical link between SET evaluations and student grades under the assumption that grades reflect learning. Some researchers have found evidence that this may be a poor assumption, and that the evaluation-grade link may reflect grade leniency to garner higher SET scores (Johnson, 2003; Greenwald & Gillmore, 1997b) or the use of SET scores as award (or punishment) for good (or bad) grades (Clayson et al, 2005; Clayson, 2004). Suggestions for improving the measure of learning include using average class grades rather than individual student grades (Marsh & Roche, 2000; Cohen, 1981;); using measures from common tests across multiple sections (Williams & Ceci, 1997; Cohen, 1981); measuring learning as the change in grades between pre- and post-tests (Hake, 2002); and measuring learning in a prerequisite class by the performance in future classes (Weinberg et al., 2007; Johnson, 2003).

The third set of problems is related to the relationship between rigor and the evaluations. If challenging courses and the associated effort required to do well are related to learning, and learning is related to good teaching, then as course rigor and effort increase, SET scores should increase as well (Greenwald & Gillmore, 1997a). However, numerous scholars have found negative associations between rigor and evaluations (Weinberg et al., 2007; Centra, 2003; Johnson, 2003). Clayson (2009) offered five explanations for the negative associations between rigor and evaluations. First, the results may be a methodological artifact, and the relationship between rigor and SETs may depend on when it is measured. When rigor was measured during the course's term, it was negatively related to SETs. When the measurement was made after the students had completed the course, the association became positive (Clayson, 2009). Second, rigor-related data has both linear and curvilinear properties, suggesting that research findings may vary depending on the level of rigor (Marsh & Roche, 2000). Students appear to give the highest evaluations to rigor that they perceive as being appropriate (Paswan & Young, 2002). Third, student self-selection into courses taught leniently may skew the statistical effect associated with rigor (Wilhelm, 2004; Johnson, 2003). Fourth, Clayson (2009) states that the perception of rigor and learning may be recursive. “If students believe that they have learned well, then rigor would be perceived as being at an appropriate level. If perceived learning was low, then rigor, at whatever level, may be perceived as being inappropriate” (Bacon and Novotny, 2002; Marks, 2000; Clayson and Haley, 1990). Fifth, the differences in how is education viewed may confound examination. Researchers and student expectations of the relationships under study may not be compatible (Chonko et al., 2002; Boex, 2000).

2.1. Research supporting SET validity

One of the early influential meta-analyses and reviews of the literature, Cohen (1981) claimed that previous research about the validity of SETs faced four major complexities. First, investigators are using different methods to measure students’ achievement. Second, the research unit varies across different papers. Some papers use “student” as the unit of analysis, but some papers focus on instructor or course level instead. Third, the student evaluation instruments are measuring multiple dimensions. Fourth, there is large variety in study settings.

In order to get a more generalized conclusion about the validity of SETs, Cohen (1981) performed a meta-analysis of 41 studies with 68 multi-section samples. The selected papers for the meta-analysis had to meet three criteria. First, the papers had to use data from actual college classes. Second, the research unit had to be instructor or class rather than student. Third, data had to have been collected from multi-section classes with a common achievement measure. Among these papers, Cohen (1981) found that positive correlations between overall course ratings and student achievement exists in 20 out of 22 multi-section courses, and 11 of them were statistically significant. Across 67 sections, the average instructor rating-student achievement correlation was 0.43, and the average course rating-student achievement correlation was 0.47. Importantly, 59 out of these 67 section correlations were positive with 30 statistically significant results. Cohen (1981) concluded that this meta-analysis provides strong support for the validity of SET ratings to measure teaching effectiveness.

Benton et al (2012) reviewed studies that compare students' SET ratings with the evaluation of some other criteria (instructor self-ratings, ratings by administrators, ratings by colleagues, ratings by alumni, ratings by trained observers and student written comments). Based on the papers they collected, student ratings were significantly and consistently related to other evaluation criteria. Benton et al (2012) stated that there are several common misconceptions about student ratings: 1) student ratings are not reliable, consistent and valid; 2) students do not appreciate good teaching while they are in school, 3) student feedbacks are not useful for teaching quality improvement, and 4) emphasis on student ratings has led to grade inflation. Their review found that, even though there is no single perfect method to measure learning, students who gave higher SET scores tend to learn more from that class in general. In summary, they found that student ratings are relatively valid and unaffected by bias, stating that while it is a useful source of data, researchers still need other information to properly evaluate faculty.

2.2. Research with mixed results on SET validity

Marsh (1997) states that SET users should consider the multidimensionality of its information while using it. Overall or global questions are not adequate because they are susceptible to context, mood and other potential biases. In particular, the global questions of SETs can only reflect the general satisfaction of students about an instructor or course. Colleges and universities should consider different teaching dimensions separately from the SET results rather than lump all the dimensions into a “puree”. Moreover, Marsh (1997) states that the SETs can only reflect students’ satisfaction about an instructor who teaches a course rather than a course that is taught, because the correlation between overall ratings of different instructors teaching the same course is much lower than correlation between overall ratings of different courses taught by the same instructor.

Stehle et al (2012) found that measuring students’ learning in different ways can produce different results using data from 883 students enrolled in 32 sections of the same medical major course. Before the classes began, the authors collected the selected students’ prior grades as a pre-test measurement. After a series of modules, the students were given an objective structured clinical examination (OSCE) that exams students’ practical skills and a multiple-choice (MC) test which exams students’ factual knowledge. In doing so, this paper measured student learning via two ways. Students were also given SETs and provided feedbacks of overall instructor quality, course quality and their subjective learning. Correlations between the objective learning measures and SET scores found that the student learning using the OSCE test results are correlated with teaching and course ratings. However, such correlations could not be found with MC grades. Based on the relation between MC test and subjective learning, researchers think that students might evaluate their learning based on the experiences they remember rather than the actual knowledge they have gained, because the OSCE might be more impressive in students' memories. An alternative explanation to the difference in correlation results is that because the OSCE measures the practical skills, students give their SET feedback only based on how much practical knowledge they can gain rather than both practical and factual knowledge.

Many of the SET research results of validity are very mixed. Some researchers suggest that alternative methods to evaluate teaching quality are required. Centra et al (1994) suggests using teaching portfolios as a tool to assess the faculty performance. In this way, academic leaders would evaluate faculty by analyzing his or her portfolio, which include perspectives of teaching effectiveness, public service, personal credentials, and other professional activities. To evaluate the validity of this method, this paper examined the difference between various groups (a dean, two faculty peers) analyzing faculty portfolio, and found general consistency in the result, suggesting that faculty portfolios can be a useful tool for faculty assessment.

2.3. Research rejecting SET validity

However, there are also some critics about the validity of SETs. Uttl et al (2017) reviewed previous meta-analyses, and found that prior meta-analysis studies suffer from serious problems. First, they claim that prior meta-analysis studies were not replicable since it is hard to locate those reference papers by the criteria listed. Second, they have issues of small study size effects. They claim that SET/learning correlations drop to nearly zero if small study size effects are adjusted. Third, they have not considered the presence of outliers in previous data.

Hoping to solve these issues, Uttl et al (2017) reran a meta-analysis on 51 up-to-date articles with 97 multi-section courses. They found that multi-section studies typically have very limited sections, inducing small study size effects. As the study size increases, the effects from outliers will decrease or even drop to zero. In addition, no matter if students' prior capacity is controlled or not, the correlations between SET and learning/achievement are very weak. Generally, this research does not support that students learning is higher from instructors with higher SET ratings. Uttl et al (2017) suggest that people in related areas at colleges or universities should set the weight of the SET use properly according to what kind of decisions are to be made. If colleges or universities want to focus on student learning, they should give a low or even no weight to SET results during the evaluation process. However, if they are simply looking to understand student satisfaction, SETs are useful.

Hobson et al (2001) summarized the characteristics of students' evaluation of teaching from the perspective of its primary application, accuracy, validity, and the developments of SET instruments. This paper explains that SETs can function as both a formative and a summative tool in the college evaluation system. Hobson et al (2001) stated that understanding how the university is developing its SET instruments and improving the reliability is important for those who are using or researching the SET survey. In addition, it would also be helpful for those who do research for this area to be aware of what roles the SET results play in the decision-making process and promotion of the college. Furthermore, he suggests that faculty should develop their own midterm evaluations in addition to their universities' official SETs so that they can adjust their courses or behaviors timely according to the evaluation feedbacks.

Love et al (2010) reviewed the data of student scores of several American colleges and tried to understand the main factors that induce grade inflation. It turns out that the more SETs are used in personnel considerations, instructors may be more inclined to engage in grade inflation to bolster SET scores. With a focus on how institutional policies affect the behaviors of faculty and students, Love et al (2010) set up a utility maximization problem to examine how faculty respond to the institutional use of SETs. This model as their theoretical basis for grade inflation and comparative static results show how SETs may diminish faculty and student effort and exacerbate grade inflation. Love et al (2010) concluded that grade targets may be an effective policy to limit grade inflation and set expectations to improve teaching and research productivity without affecting student effort.

Researchers have found that SETs are significantly and consistently related to the ratings of alumni, student achievement in multi-section validity studies, and faculty self-evaluations and trained evaluators, but not related to colleague and administrator ratings. Generally, researchers believe that SETs should be designed and used appropriately for summative feedback, and that college administrators should use faculty portfolios or other methods for formative evaluations of teaching.

3. Model and Hypothesis

Our hypothesis is that SET feedback captures teaching effectiveness in the form of learning. In other words, the more students learn, the higher are the instructor and course SET scores. This paper will test this hypothesis using two methods—one using individual students as the unit of analysis, and one using course sections as the unit of analysis. First, we use a probit regression model to examine the impact of students' self-reported learning results on SET scores.[1] While this first method uses a subjective measure of learning to validate teaching effectiveness, the second method uses a multi-level regression model designed by Weinberg et al (2009). This second approach attempts to develop an objective measure of learning based on the link between performance in prerequisite courses (e.g., ECON 101, PSY 101) and subsequent courses (e.g., ECON 201, PSY 201). We present each approach in turn.

[1] Because we have over 200,000 observations, we apply a bootstrap method using 1,000 replications of the regression on 1,000-observation samples to obtain coefficient estimates and standard errors.

3.1. Probit regression

Our first model is shown in equation 1. The SET score, $\boldsymbol{Y_{SET}}$, is the dependent variable, and $\boldsymbol{X_{Learn}}$ denotes our subjective measure of learning, which is the students' response to a question on the SET survey asking: *Overall, how much do you feel you've learned in this course?* A value of 1 represents responses "more than most courses" or "much more than most courses," and 0 otherwise. The dependent variable is a binary variable that equals 1 when the SET score is a 4 or 5 (the highest levels on scale of five), and 0 otherwise. A grade variable, $\boldsymbol{X_{Grade}}$, is also included in the model; it is a binary variable that equals 1 when the grade is an A, and 0 otherwise. Our model also includes $\boldsymbol{D_{Class}}$, $\boldsymbol{D_{Instructor}}$, and $\boldsymbol{D_{Student}}$, which denote three sets of control variables for the class, student, and instructor backgrounds.

$$\boldsymbol{Y_{SET} = \beta_1 \cdot X_{Learn} + \beta_2 \cdot X_{Grade} + \beta_3 \cdot D_{Class} + \beta_4 \cdot D_{Instructor} + \beta_5 \cdot D_{Student} + \varepsilon_1} \quad [1]$$

Support for the validity of SETs as a measure of teaching effective is found when the estimated coefficient, β_1, is positive and statistically significant. This finding would suggest that controlling for grades, student, class, and instructor characteristics, the students' belief of having learned more than in most classes leads them to rate their instructor and courses more favorably.

This model is able to address the grade leniency issue by using grades as controls and using the measure of learning to test the teaching effectiveness hypothesis, but our measure of learning is subjective. This model is also able to control for rigor as one of the class characteristics. We measure rigor subjectively using student responses to the following SET survey question: *For me, the level of difficulty of the course content was…".* Rigor equals 1 when students complete the question sentence with "much more than most courses" or "more than most courses," and 0 otherwise. However, this first model does not address the limits of using individual level of analysis, the situational nature of the data, nor use an objective measure of learning.

3.2. Multi-level regression

Our second method uses a multi-level regression model designed by Weinberg et al (2009). This approach attempts to develop an objective measure of learning based on the link between performance in prerequisite and subsequent courses. This innovative approach selects a sample of students who took classes in paired course sequences and observes how their performance in the base course (the prerequisite) affects their performance in the following courses for which the prerequisite was required. While Weinberg et al (2009) limited their study to micro- and macro-economics courses, this paper applies this method to base-subsequent course pairs for a number of different subjects: Accounting, Biology, Chemistry, Communication, Economics, and Psychology.[2]

The first two levels of our multi-level regression model focuses on students as the unit of analysis. The first level is shown in equation 2, which regresses—by section—student grades in the base course, $\boldsymbol{Y_{Grade}}$, on $\boldsymbol{D_{Base}}$, a vector of dummy variables marking which base course this student took, an $\boldsymbol{D_{Student}}$, a series of control variables for students' background, including gender, race and SAT grades. We are interested in the estimated coefficient, $\boldsymbol{\theta_1}$, which is a vector of average base course grades by section after controlling for student characteristics.

$$\boldsymbol{Y_{Grade} = \theta_1 \cdot D_{Base} + \theta_2 \cdot D_{Student} + \varepsilon_2} \qquad [2]$$

[2] These subjects were chosen because they have a large number of sections, which is an important requirement for our second method.

The second level of this model, shown in equation 3, also uses students as the unit of analysis. It regresses—by section—student grades in the subsequent course, $\boldsymbol{Y'_{Grade}}$, on $\boldsymbol{D_{Subsequent}}$, the vector of dummy variables marking the subsequent courses taken by students who took the base courses in the first level equation, and $\boldsymbol{D_{Base}}$ and $\boldsymbol{D_{Student}}$, the same vectors used in equation 2 controlling for base dummies student characteristics. Here, we are interested in the estimated coefficient, $\boldsymbol{\theta_9}$, which is a vector of average subsequent course grades by section attributable to the base course. That is, $\boldsymbol{\theta_9}$ captures the effect of base courses on subsequent courses in paired course sequences, after controlling for the subsequent courses' own effects and student characteristics. This learning vector, $\boldsymbol{\theta_4}$, is our objective measure of student learning.

$$\boldsymbol{Y'_{Grade} = \theta_3 \cdot D_{Subsequent} + \theta_4 \cdot D_{Base} + \theta_5 \cdot D_{Student} + \varepsilon_3} \quad [3]$$

At this point, each section of a base course will have an average base grade vector ($\boldsymbol{\theta_1}$) and an average base learning vector ($\boldsymbol{\theta_4}$). These vectors become variables in the third stage of our multi-level regression model, which focuses on base course sections as the unit of analysis. Shown in equations 5 and 6, respectively, we regress the average SET scores by section, $\boldsymbol{Y_{AvgTeach}}$ and $\boldsymbol{Y_{AvgCourse}}$, on $\boldsymbol{\theta_1}$, the average base grade vector; $\boldsymbol{\theta_4}$, the average base learning vector; and $\boldsymbol{D_{Instructor}}$, a set of three instructor control variables, including instructors' ethnicity, tenure position and gender. Finally, we are interested in the estimated coefficients, $\boldsymbol{\beta_7}$ and $\boldsymbol{\beta_{10}}$, which are the effects of student learning on instructor and course ratings, respectively.

$$\boldsymbol{Y_{AvgTeach} = \beta_6 \cdot \theta_1 + \beta_7 \cdot \theta_4 + \beta_8 \cdot D_{Instructor} + \varepsilon_4} \quad [4]$$

$$\boldsymbol{Y_{AvgCourse} = \beta_9 \cdot \theta_1 + \beta_{10} \cdot \theta_4 + \beta_{11} \cdot D_{Instructor} + \varepsilon_5} \quad [5]$$

This represents our strongest validity test for the effectiveness of teaching. We are using objective measures of student learning, course sections as the ultimate unit of analysis, separate subject-based analyses to address the situational nature of the data, and includes controls for grade leniency. However, unlike the first model, this second model does not include a control for course rigor.

4. Data

The data used in this paper comes from the administrative database of a large public research university. The following data were obtained from Fall 2013 to Spring 2017 semesters: 1) student grades and characteristics, 2) student evaluations of instructors and courses (i.e., SETs), and 3) faculty information. The student course and grade information across four academic years was combined with students' personal information, such ethnicity, gender, and SAT verbal and math scores. Then, the student information is combined with course and instructor information through the students' SET database.

Tables 1a and 1b show the summary statistics of our two research samples. In Table 1a, we summarize data for 207,927 student evaluations, including SET scores, measures of learning, and student, instructor, and class characteristics. Student characteristics include demographic information, student grades, grade expectations, and whether students thought the course was hard or interesting, and whether the students missed classes and studied a lot. Instructor characteristics include ethnicity, gender, tenure-status, and years of experience. Class characteristics include class size, whether it was a STEM course, and whether course was taken on the university's main campus.

In Table 1a, we see that the average SET scores for instructors and courses are 3.93 and 3.71, respectively, and that just more than half of the students in our sample reported learning more than most courses (54.8%) and obtaining a grade of A (57.1%). Just less than half of the instructors were tenured or on tenure-track (44.8%) and female (45.1%), while a quarter of the instructors are ethnic minorities (24.9%). The average instructor experience is about nine years. In terms of student characteristics, females (56.2%) and ethnic minorities (27.1%) are represented in similar percentages as instructors, nearly half rated their course as being more interesting than other courses (47.6%) and just over a third claimed that their course is harder than their other courses (36.6%).

In Table 1b, we summarize our data for 78 different course sequences among six subject departments: Accounting, Biology, Chemistry, Communication, Economics, and Psychology.[3] Interestingly, in this sample of evaluations, females were the clear majority in Biology (67.9%), Chemistry (60.1%), and Psychology (74.6%). Proportions of ethnic minorities ranged from 38% (Accounting and Microeconomics) to 53.3% (Chemistry). Those with the highest SAT scores tended to take Accounting, while those with the lowest tended to take Psychology. Average grades in Economics and Accounting tended to be higher than other subjects, while they were lowest in Chemistry. Table 1b also shows that the average GPA for students in this sample range from 2.73 to 3.39.[4] The average teaching evaluation scores range from 3.76 to 4.12, while the average course evaluation scores range from 3.47 to 4.0, suggesting that students may be tougher critics in evaluating courses than instructors.

5. Results

We test our hypothesis first using ordered probit regressions to examine the relationship between student evaluations and our subjective measure of learning; the dependent variables are

[3] The course sequences for each subject are listed in the Appendix.
[4] Numerical grades are obtained as follows: A = 4.0, A- = 3.67, B+ = 3.33, B = 3.0, B- = 2.67, C+ = 2.33, C = 2.0, C- = 1.67, D+ = 1.33, D = 1.0, F = 0.

the instructor and course evaluations, which are 5-item Likert-scaled student responses –ranging from Poor (=1) to Excellent (= 5)—to the question, "*What is your overall rating of the instructor's teaching/course*." Because we have a very large number of observations (207,927), we use a bootstrapping technique to resample the data for appropriate estimates of standard errors. Using sample sizes of 1,000 evaluations each iteration, the bootstrapping procedure repeats the regression 1,000 times, estimating our model coefficients and robust standard errors clustered within instructors and courses.

Table 2 shows that there are positive and statistically significant relationships between our subjective measures of learning (*More Learning* and *Less Learning*) and overall SET scores for instructors and courses. This is true even after controlling for grades, which are also shown to have positive and statistically significant relationships with overall SET scores for instructors and courses. Thus, if students feel that they have learned more (less) from the class than other classes they have taken, they tend to give higher (lower) evaluation scores, even after controlling for grades, expected grades, and other student, class, and instructor characteristics. With our models explaining about a quarter of the variation in the dependent variables, this is modest evidence in support of SETs as valid measures of teaching effectiveness.

Table 2 also reveals bias in how students evaluate minority and older instructors and courses taught by minority instructors. Students tend to evaluate minority instructors and the courses they teach lower than white instructors and the courses they teach. This finding suggests caution in using SETs for faculty promotion, tenure, and reappointments. Moreover, we find that more experienced instructors received lower ratings, and female students may be tougher in evaluating instructors than male students, but these two patterns do not apply to course evaluations.

In terms of our measure of rigor (*More Hard*, *Less Hard*), we find similar evidence as others in the literature; students downgrade (upgrade) evaluations of instructors and courses when they believe their courses are harder (easier) than others they take, controlling for instructor, class, and student factors. Other factors were estimated as statistically significant, including class size (bigger classes are related to higher SET ratings); higher cumulative GPA (better performing students grade instructors and courses lower, on average); students with high (low) grade expectations rate instructors and courses higher (lower); and more (less) interest in the course subject lead to higher (lower) ratings, controlling for other factors. Interestingly, students who miss time in class rate instructors and courses lower to the extent that they miss classes, as the estimated coefficients on the *Missed Lots of Time* variable is twice the magnitude for *Missed Some Time*. Also interesting, we find that transfer students tend to give higher course evaluations than non-transfer students, and this difference does not apply to instructors, suggesting an appreciation for the course relative to those from where they transferred.

While these results provide modest support for the validity of SETs as measures of teaching effectiveness, they also support the cry for caution in using SETs for high stakes personnel decisions due to the significant findings of bias.

However, numerous researchers have criticized subjective assessments of learning. Prior studies have shown that students have limited or biased impressions about their class experience and learning (Stehle et al, 2012), so we apply the method introduced by Weinberg et al's (2009) to measure learning more objectively. We turn now to the findings using the 3-stage Weinberg approach.

Table 3 summarizes the model fit results from the first two stages (equations 2 and 3) of the multi-level analyses. The dependent variable in these stages are student grades, and they range from 78 observations in Biology to 1,008 in Accounting. With a few exceptions, we are generally satisfied with the requisite number of observations, as the variation in the dependent variables are reasonably well-explained by the models. The Adjusted R^2's range from 18% (Accounting) to 74% (Microeconomics) in the first stage, and from 3% (Chemistry) to 80% (Biology) in the second stage. The relative small sample size in Biology (n = 78) did not seem to pose a problem with model fit, while the sample size in Chemistry (n = 321) seemed inadequate in the second stage.

These 1st and 2nd stage regressions include student characteristics and class section dummy variables as the independent variables, and the lower panel in Table 2 shows their direction of signs and level of statistical significance. In general, the estimates of the impact of rigor (*More Hard)* show that students who deem their instructors/courses as harder than others they have taken make lower grades. We found no gender effects except with Biology, Communications, and Micro- and Macroeconomics, where female instructors give higher grades than male instructors. The effects of race/ethnicity take mixed patterns, some positive, and others negative, while SAT scores are generally positively related to grades.

From the first and second stages of the multi-level regression model, we obtain estimated coefficients representing the average course grade ($\boldsymbol{\theta_1}$) and average learning ($\boldsymbol{\theta_4}$). These coefficients are at the course section level, and now can enter into an analysis in the third stage at the instructor/course section level. Tables 4 and 5 present the results of third stage analyses for instructors and courses, respectively, where our dependent variables are the instructor and course evaluations.

Table 4 shows that only student learning ($\boldsymbol{\theta_4}$) in Biology is positively related to instructor evaluations, but that student learning in Accounting is negatively related to instructor evaluations. We find no support for SET validity in the other subjects, except in Psychology (probability of statistical chance or error = .11) the level of statistical significance fell just short of normally acceptable levels.

In terms of the other impacts on instructor evaluations, average course grades ($\boldsymbol{\theta_1}$) are found to have a positive impact in Biology, Microeconomics, and Psychology. Mixed results are found for instructor rank and gender; higher ranked instructors received lower rankings in Chemistry and Microeconomics, and higher rankings in Psychology. Female instructors received lower rankings in Accounting, Communication, and Macroeconomics, but higher rankings in Microeconomics and Psychology. Minority instructors received higher rankings in Accounting, and lower rankings in Psychology.

Table 5 shows similar findings for course evaluations—only in Biology do we find a positive link between student learning and course evaluations. Similar with instructors, average course grades are found to have a positive impact on course evaluations in Biology, Microeconomics, and Psychology, and a similar pattern of mixed results are found in the impacts of instructor rank and gender. Courses taught by higher ranked instructors received lower rankings in Chemistry and Microeconomics, and higher rankings in Psychology. Courses taught by female instructors received higher rankings in Microeconomics and Psychology, and courses taught by minority instructors received lower rankings in Psychology.

The analyses above addressed the three types of problems that confound research on the relationship between learning and teaching evaluations. The situational problems were addressed by conducting the multi-level analyses by discipline, and we indeed found different findings for different disciplines. The methodological problems of measuring learning and controlling for grade leniency were addressed in three ways. First, we used average class grades in our multi-level analysis, and found effects consistent with using individual student grades in our ordered probit analysis. Second, we controlled for subject in our multi-level analyses as a proxy for using common tests across multiple sections. Third, we did not have access to grade measures from pre- and post-tests, as suggested by Hake (2002), but our multi-level analyses did use learning in a prerequisite class by the performance in future classes, as demonstrated by Weinberg et al. (2009). We only found a learning-evaluation link in Biology from our multi-level analyses. Finally, the problem of controlling for rigor was addressed, and we found similar evidence as others in the literature, that rigor is negatively associated with evaluations (Weinberg et al., 2007; Centra, 2003; Johnson, 2003).

Our analyses summarized in Table 2 (ordered probit regressions) and Tables 4 and 5 (multi-level regressions) find similar results with regard to impacts of grades and rigor on instructor and course evaluations. Both sets of analyses find broad evidence that after controlling for instructor, class, and student characteristics, grades are positively related to instructor and course evaluations, and instructor and course rigor are negatively related. However, our findings were indeed situational to both the discipline and analytic method. Using ordered probit regression analyses on pooled data making no distinction between disciplines and measuring learning subjectively, we found reasonably strong evidence of a small link between learning and evaluations. Using the multi-level regressions that do control for discipline and measure learning objectively, we only find a modest positive learning-evaluation link in Biology, with a similar effect found in Psychology with less confidence.

6. Conclusion

The answer to the key question, "are the evaluations that students make of courses and instructors related to student learning?", seems to depend the method of analysis and measure of learning. We find results consistent with Clayson's (2009) conclusion that whatever relationship exists between learning and teaching evaluations, it is “situational and not applicable to all teachers, academic disciplines, or levels of instruction.” Also consistent with Clayson’s conclusions, we find that the more objectively learning is measured, the less likely it is to be related to the evaluations.

When we used a subjective measure of learning, we found support for our hypothesis of a positive relationship between student learning and evaluations. However, when we used an objective measure of learning, we found such support in one, maybe two disciplines. Thus, universities should give use SETs with caution, giving proper weight to the results of student evaluations, which may be biased negatively toward older, female, and/or minority instructors, and may punish instructors for delivering rigorous instruction. They should not be used to impact personnel decisions, but SETs may still be a useful tool to measure students satisfaction with their instructors and courses. Teaching effectiveness has numerous dimensions, and a similar analysis can be applied to individual dimensions of teaching effectiveness to determine which elements are captured by SETs. While students may be whimsical and capricious at times in their evaluations of their teachers and courses, but if they say they are learning, that might not be so bad.

References

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of educational Research*, *51*(3), 281-309.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential baises, and utility. *Journal of educational psychology*, *76*(5), 707.

Weinberg, B. A., Hashimoto, M., & Fleisher, B. M. (2009). Evaluating teaching in higher education. *The Journal of Economic Education*, *40*(3), 227-261.

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. American psychologist, 52(11), 1187.

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. Studies in Educational Evaluation, 54, 22-42.

Stehle, S., Spinath, B., & Kadmon, M. (2012). Measuring teaching effectiveness: Correspondence between students' evaluations of teaching and different measures of student learning. Research in Higher Education, 53(8), 888-904.

Love, D. A., & Kotchen, M. J. (2010). Grades, course evaluations, and academic incentives. Eastern Economic Journal, 36(2), 151-163.

Hobson, S. M., & Talbot, D. M. (2001). Understanding student evaluations: What all faculty should know. College teaching, 49(1), 26-31.

Centra, J. A. (1994). The use of the teaching portfolio and student evaluations for summative evaluation. The Journal of Higher Education, 65(5), 555-570.

Benton, S. L., & Cashin, W. E. (2012). Student ratings of teaching: A summary of research and literature (IDEA Paper No. 50). Manhattan, KS: The IDEA Center. Retrieved April 8, 2012.

Table 1a. Summary Statistics for Probit Regression Data

N = 207,927	Average	Standard Deviation	Minimum	Maximum
SET overall score				
Instructor	3.93	1.13	1	5
Course	3.71	1.11	1	5
Measures of learning				
More Learning	.548	.498	0	1
Less Learning	.120	.325	0	1
GradeA	.571	.495	0	1
Instructor characteristics			0	1
Tenured/Tenure Track	.448	.497	0	1
Faculty Female	.451	.498	0	1
Faculty Minority	.249	.432	0	1
Experience	8.98	9.58	0	59
Class characteristics				
Class Size	58.1	62.5	1	403
Non-STEM	.641	.480	0	1
Main campus	.791	.406	0	1
Student characteristics				
Major Required	.552	.497	0	1
Upper class	.594	.491	0	1
Minority	.271	.445	0	1
High Cumulative GPA	.370	.483	0	1
Expected A	.570	.495	0	1
Expected DF	.011	.103	0	1
Female	.562	.496	0	1
Transfer	.167	.373	0	1
Missed Some Time	.178	.382	0	1
Missed Lots of Time	.084	.278	0	1
Study None	.056	.231	0	1
Study Lots	.098	.297	0	1
More Interest	.476	.499	0	1
Less Interest	.175	.380	0	1
More Hard	.366	.482	0	1
Less Hard	.221	.415	0	1

Table 1b. Summary Statistics for Multi-Level Regression Data

	Accounting	Biology	Chemistry	Communication	Microeconomics	Macroeconomics	Psychology
Gender							
Female	52.2%	67.9%	60.1%	6.5%	35.8%	37.6%	74.6%
Race/Ethnicity							
Asian American	8.9%	6.4%	25.9%	4.0%	5.1%	4.4%	10.3%
Black	2.0%	9.0%	7.5%	6.7%	3.9%	8.8%	7.8%
Latino	10.0%	12.8%	10.0%	20.3%	9.3%	7.3%	18.8%
Other	12.6%	14.1%	6.2%	10.7%	12.5%	9.9%	6.6%
NRA	4.7%	3.8%	3.7%	4.3%	7.4%	9.9%	0.6%
SAT							
Math	618.5	519.7	558.4	539.4	579.1	586.4	497.5
(St.Dev., Min, Max)	(78.9, 400, 800)	(71.6, 400, 690)	(82.3, 350, 750)	(78.8, 350, 740)	(104.1, 370, 800)	(103.7, 370, 800)	(69.9, 280, 680)
Verbal	574.1	530.8	528.4	522.3	539.9	548.2	499.6
(St.Dev., Min, Max)	(76.5, 360, 780)	(67.0, 410, 740)	(77.8, 350, 730)	(80.1, 360, 800)	(99.6, 320, 760)	(92.3, 340, 800)	(67.6, 280, 690)
1st & 2nd Stage Summary							
Base Course Grade	3.35	3.00	2.92	3.19	3.39	3.37	3.34
(St.Dev., Min, Max)	(0.77, 0, 4)	(0.94, 0.7, 4)	(0.87, 0, 4)	(0.65, 0.7, 4)	(0.62, 0.7, 4)	(0.61, 1.7, 4)	(0.75, 0, 4)
Subsequent Course Grade	3.33	2.94	2.73	3.11	3.23	3.22	3.38
(St.Dev., Min, Max)	(0.72, 0, 4)	(0.93, 1, 4)	(1.03, 0, 4)	(0.68, 1, 4)	(0.85, 0, 4)	(0.82, 0, 4)	(0.76, 0, 4)
3rd Stage Summary							
Number of Sections	37	39	57	71	67	57	78
Avg SET Teaching Scores	4.12	3.76	3.82	3.80	3.87	4.08	4.07
Avg SET Course Scores	3.90	3.65	3.47	3.73	3.79	4.00	3.90
% Instructors Minority	11%	10%	33%	32%	24%	21%	6%
% Instructors Female	16%	21%	63%	34%	3%	2%	6%
% Instructors Tenure/ Tenure Track	41%	23%	7%	9%	27%	14%	8%

Table 2 – Bootstrapped Probit Regression Results

SET Instructor Ratings			SET Course Ratings		
	Coeff.	**Std. Err.**		**Coeff.**	**Std. Err.**
Measures of learning			**Measures of learning**		
Grade A	0.12***	0.03	Grade A	0.14***	0.02
More Learning	0.91***	0.03	More Learning	0.92***	0.02
Less Learning	-0.95***	0.04	Less Learning	-0.93***	0.04
Instructor characteristics			**Instructor characteristics**		
Tenured/Tenure Track	-0.10	0.06	Tenured/Tenure Track	-0.05	0.04
Faculty Female	-0.02	0.05	Faculty Female	0.00	0.04
Faculty Minority	-0.16***	0.06	Faculty Minority	-0.12**	0.05
Experience	-0.01**	0.00	Faculty Experience	0.00	0.00
Class characteristics			**Class characteristics**		
Class Size	0.00**	0.00	Class Size	0.00**	0.00
Class Size-square	0.00	0.00	Class Size-square	0.00*	0.00
Non-STEM	0.00	0.03	Non-STEM	0.02	0.02
Main campus	-0.06	0.04	Main campus	-0.09***	0.03
Student characteristics			**Student characteristics**		
Major Required	-0.09**	0.04	Major Required	-0.08***	0.03
Upper class	0.05	0.04	Upper class	0.04	0.03
Minority	0.02	0.02	Minority	0.05**	0.02
High Cumulative GPA	-0.07***	0.02	High Cumulative GPA	-0.09***	0.02
Expected A	0.25***	0.03	Expected A	0.32***	0.02
Expected DF	-0.16*	0.09	Expected DF	-0.28**	0.11
Female	-0.07**	0.03	Female	-0.01	0.02
Transfer	0.04	0.03	Transfer	0.05**	0.02
Missed Some Time	-0.06**	0.03	Missed Some Time	-0.08***	0.03
Missed Lots of Time	-0.16***	0.04	Missed Lots of Time	-0.15***	0.04
Study None	-0.07	0.05	Study None	-0.06	0.05
Study Lots	-0.08*	0.04	Study Lots	-0.05	0.04
More Interest	0.22***	0.03	More Interest	0.39***	0.02
Less Interest	-0.14***	0.04	Less Interest	-0.32***	0.03
More Hard	-0.36***	0.03	More Hard	-0.43***	0.03
Less Hard	0.23***	0.04	Less Hard	0.21***	0.03
_constant	0.46	0.09	_constant	-0.03	0.07
Number of observations		207,927	Number of observations		207,927
Sample sizes		1,000	Sample sizes		1,000
Replications		1,000	Replications		1,000
Clusters in Course and Instructor		6,744	Clusters in Course and Instructor		6,744
Wald chi-square		3,807	Wald chi-square		5,596
Pseudo R-squared		0.243	Pseudo R-squared		0.268

Note: * p<0.1; ** p<0.05; *** p<0.01

Table 3 - 1st and 2nd Stage Regression Results

	Accounting		**Biology**		**Chemistry**		**Communication**		**Microeconomics**		**Macroeconomics**		**Psychology**	
	1st Stage	2nd Stage	1st Stage	2nd Stage	1st Stage	2nd Stage	1st Stage	2nd Stage	1st Stage	2nd Stage	1st Stage	2nd Stage	1st Stage	2nd Stage
Number of Observation	1008	779	78	53	321	299	300	234	257	175	274	186	319	310
F	6.59	2.23	2.72	4.20	2.60	1.09	6.61	2.21	11.93	2.87	5.23	1.65	4.03	2.23
Prob > F	0.000	0.000	0.001	0.004	0.000	0.302	0.000	0.000	0.000	0.003	0.000	0.053	0.000	0.000
R^2	0.284	0.434	0.694	0.927	0.390	0.494	0.711	0.859	0.821	0.954	0.630	0.895	0.599	0.573
Adjusted R^2	0.241	0.239	0.438	0.706	0.240	0.040	0.603	0.470	0.752	0.622	0.509	0.354	0.450	0.316
Root MSE	0.675	0.634	0.705	0.441	0.755	1.022	0.409	0.484	0.308	0.539	0.430	0.626	0.553	0.609
Are estimated coefficients statistically significant?														
Rigor	Yes -***	Yes -***	Yes -**	No	Yes -**	Yes -**	Yes -***	No	Yes -***	Yes -***	Yes -*	No	Yes -***	No
Gender	No	No	No	Yes +**	No	No	Yes +***	No	Yes +**	No	Yes +*	No	No	No
Race	Yes -***	Yes -***	Yes +***	Yes -***	Yes +-***	No	Yes +-***	No	Yes +***	Yes +-***	Yes -*	No	Yes +**	No
SAT's	Yes +***	Yes +***	Yes +**	Yes +***	Yes +***	No	Yes +***	Yes +*** -**	Yes +***	No	Yes +***	No	Yes +***	Yes +***

Notes: Direction of sign denoted by + or –. Significance denoted by * p<0.1, ** p<0.05, or *** p<0.01.

Table 4 - Teaching Evaluation

Dependent Variable = Average Instructor SET Scores	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Subject	**Accounting**	**Biology**	**Chemistry**	**Communication**	**Microeconomics**	**Macroeconomics**	**Psychology**
Independent Variable							
Students' Learning	-.721*	.387***	.124	.020	-.077	.130	.209
	(.337)	(.056)	(.164)	(.106)	(.052)	(.186)	(.126)
Control Variable							
Course Grade	.232	.785**	.556	-.025	.564***	.184	.521***
	(.343)	(.351)	(.400)	(.230)	(.184)	(.236)	(.150)
Instructor Rank	-.064	.363	-1.016***	-.577	-.639**	-.020	.701***
	(.115)	(.361)	(.300)	(.369)	(.256)	(.287)	(.174)
Instructor Gender	-.314*	-.080	.263	-.800***	.726***	-.455*	.792***
	(.149)	(.533)	(.409)	(.273)	(.236)	(.248)	(.242)
Instructor Race	1.172**	.059	-.130	-.315	.223	.198	-.454*
	(.432)	(.363)	(.433)	(.263)	(.338)	(.292)	(.242)
Number of observations	37	39	57	71	67	57	78
R-squared	.112	.333	.108	.279	.116	.043	.271
Root MSE	1.055	.890	1.044	.797	1.178	.886	.690

Note: * p<0.1; ** p<0.05; *** p<0.01

Table 5 - Course Evaluation

Dependent Variable = Average Course SET Scores							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Subject	**Accounting**	**Biology**	**Chemistry**	**Communication**	**Microeconomics**	**Macroeconomics**	**Psychology**
Independent Variable							
Students' Learning	-.454	.389***	-.246	-.102	-.022	.093	.115
	(.433)	(.093)	(.182)	(.095)	(.058)	(.162)	(.080)
Control Variable							
Course Grade	.128	.798**	.513	.089	.580**	.068	.532***
	(.275)	(.335)	(.331)	(.193)	(.239)	(.229)	(.173)
Instructor Rank	-.158	.237	-.554***	-.012	-.682**	-.257	.532***
	(.144)	(.189)	(.202)	(.126)	(.261)	(.221)	(.155)
Instructor Gender	-.117	-.435	.464	-.375	.746***	-.358	.774***
	(.145)	(.545)	(.359)	(.260)	(.236)	(.270)	(.138)
Instructor Race	.461	.312	.509	-.136	.326	-.166	-.727***
	(.370)	(.192)	(.322)	(.211)	(.295)	(.294)	(.195)
Number of obs	37	39	57	71	67	57	78
R-squared	.048	.438	.159	.106	.118	.039	.266
Root MSE	.934	.800	.929	.751	1.148	.754	.699

Note: * p<0.1; ** p<0.05; *** p<0.01

Appendix – Prerequisite and Subsequent Course List

Subject Name	Course Name	Level
Biology	Principles of Biology	Prerequisite
12 Following	Principles of Biology	Subsequent
	Introduction to Biochemistry	Subsequent
	Cell Biology	Subsequent
	Human Genetics	Subsequent
	Genetics	Subsequent
	Fundamentals of Microbiology	Subsequent
	Introduction to Molecular Evolution and Bioinformatics	Subsequent
	Molecular Biology and Genetics of Prokaryotes	Subsequent
	Research Literature in Molecular and Cell Biology	Subsequent
	Structure and Function of Biological Macromolecules	Subsequent
	Advanced Biochemistry Laboratory	Subsequent
	Basic Immunology	Subsequent
Accounting	Principles of Financial Accounting	Prerequisite
49 Following	Principles of Managerial Accounting	Subsequent
	Introduction to a Profession	Subsequent
	Intermediate Accounting I	Subsequent
	Intermediate Accounting II	Subsequent
	Cost Accounting	Subsequent
	Federal Income Taxes	Subsequent
	Advanced Accounting	Subsequent
	Financial Statement Analysis and Business Valuation	Subsequent
	Assurance Services	Subsequent
	Taxation of Business Entities	Subsequent
	Principles of Managerial Accounting	Subsequent
	Financial Management	Subsequent
	Managerial and Interpersonal Behavior	Subsequent
	Introduction to Marketing Management	Subsequent
	Effective Business Writing	Subsequent
	Financial Management	Subsequent
	Investments and Security Analysis	Subsequent
	Principles of Investments and Derivatives	Subsequent
	Real Estate Investments	Subsequent
	Real Estate Finance	Subsequent
	Applications in Financial Management	Subsequent
	Advanced Issues in Security Valuation	Subsequent
	Fixed Income Securities	Subsequent
	Advanced Issues in Asset Allocation and Portfolio Management	Subsequent

	Financial Derivatives and Risk Management	Subsequent
	Global Financial Management	Subsequent
	Financial Services	Subsequent
	Entrepreneurial Finance	Subsequent
	Security Valuation and Portfolio Management	Subsequent
	Alternative Investments and Risk Management	Subsequent
	Managerial and Interpersonal Behavior	Subsequent
	International Business	Subsequent
	Opportunity Generation, Assessment, and Promotion	Subsequent
	Venture Planning, Management, and Growth	Subsequent
	Managerial Negotiations	Subsequent
	Strategy, Policy and Planning	Subsequent
	Strategic Analysis	Subsequent
	Introduction to Marketing Management	Subsequent
	Consumer Behavior	Subsequent
	Marketing Research	Subsequent
	Marketing Planning and Strategy	Subsequent
	Global Marketing Strategy	Subsequent
	Professional Selling	Subsequent
	Sales Management and Leadership	Subsequent
	Integrated Marketing Communications in the Digital Age	Subsequent
	Marketing and Digital Analytics	Subsequent
	Digital Marketing	Subsequent
	Strategic Brand Management	Subsequent
Chemistry	General Chemistry	Prerequisite
10 Following	General Chemistry	Subsequent
	The Science of Chemistry	Subsequent
	Organic Chemistry	Subsequent
	Organic Chemistry	Subsequent
	Organic Chemistry	Subsequent
	Organic Chemistry Laboratory	Subsequent
	Organic Chemistry Laboratory	Subsequent
	Descriptive Inorganic Chemistry	Subsequent
	Physical Chemistry	Subsequent
	Physical Chemistry	Subsequent
Communication	The Process of Communication	Prerequisite
11 Following	Professional Communication	Subsequent
	Media Literacy and Criticism	Subsequent
	Fundamentals of Digital Production	Subsequent
	Research Methods in Communication	Subsequent
	Persuasion	Subsequent

	Interpersonal Communication	Subsequent
	Effects of Mass Media	Subsequent
	Gender and Communication	Subsequent
	Advanced Media Effects	Subsequent
	Children and Mass Media	Subsequent
	Communication Technology and Social Change	Subsequent
Micro	Principles of Microeconomics	Prerequisite
17 Following	Economic History of Europe	Subsequent
	Economic History of the United States	Subsequent
	Intermediate Microeconomic Theory	Subsequent
	Intermediate Macroeconomic Theory	Subsequent
	Mathematical Economics	Subsequent
	Empirical Methods in Economics I	Subsequent
	Operations Research	Subsequent
	Information Technology for Economics	Subsequent
	Money and Banking	Subsequent
	Economics of the Global Economy	Subsequent
	Women and Minorities in the Labor Market	Subsequent
	Economics of Poverty	Subsequent
	Economics of the Oceans	Subsequent
	Economic Development in Latin America and the Caribbean	Subsequent
	Transitional Economies of Russia and Eastern Europe	Subsequent
	Writing in Economics	Subsequent
	International Finance	Subsequent
Macro	Principles of Macroeconomics	Prerequisite
20 Following	Economic History of Europe	Subsequent
	Economic History of the United States	Subsequent
	History of Economic Thought	Subsequent
	Intermediate Microeconomic Theory	Subsequent
	Intermediate Macroeconomic Theory	Subsequent
	Mathematical Economics	Subsequent
	Empirical Methods in Economics I	Subsequent
	Operations Research	Subsequent
	Information Technology for Economics	Subsequent
	Money and Banking	Subsequent
	Economics of the Global Economy	Subsequent
	Women and Minorities in the Labor Market	Subsequent
	Economics of Poverty	Subsequent
	Economic Development in Latin America and the Caribbean	Subsequent
	Transitional Economies of Russia and Eastern Europe	Subsequent
	Writing in Economics	Subsequent

	International Trade	Subsequent
	Public Finance	Subsequent
	Urban and Regional Economics	Subsequent
	Economic Development	Subsequent
Psychology	General Psychology I	Prerequisite
1 Following	General Psychology II	Subsequent

Resource: https://catalog.uconn.edu/directory-of-courses/