From “Traditional” to Research Based Instructional Strategies: An Assessment of Learning Gains

By SARAH B. COSGROVE AND NEAL H. OLITSKY*

* Cosgrove and Olitsky: University of Massachusetts Dartmouth, 285 Old Westport Rd, N. Dartmouth, MA 02747 (sarah.cosgrove@umassd.edu, neal.olitsky@umassd.edu).

I. Introduction

Cognitive science literature reveals research-based instructional strategies (RBIS) that are proven to help students learn, retain, and transfer skills and knowledge (Brown, Roediger III, and McDaniel 2014). Still, many economics professors are not using these strategies. Faculty cite additional preparation time and the inability to cover all of the necessary topics if class time is devoted to active learning (Goffe and Kauper 2014).

Flipping the classroom helps to solve the latter problem. Students in a flipped course familiarize themselves with material before class. Class time is used for activities that help students achieve higher levels of learning.

We conduct an experiment to illustrate what average instructors can expect in terms of student learning gains should they change their teaching method. Using a pre/post analysis, we compare students’ learning gains in a flipped class that uses refined RBIS to the gains in a “traditional” class and estimate the differential gains in assessment scores for students in the RBIS course. Our results indicate significant, positive differential learning gains for students in the RBIS class, which are more prevalent in content presented earlier in the semester. To our knowledge, this study is the first to compare outcomes from a flipped-RBIS economics class to a traditional economics class. Moreover, in addition to traditional OLS estimation, we estimate the treatment effect using inverse probability weighting regression adjustment, providing comparable treatment and control groups, and accounting for potential selection into treatment. This method allows us to estimate the effect of participating in an RBIS course on learning gains.

II. Experiment Design

Our control group is a “typical” principles class and the treatment group employs refined RBIS. In line with Goffe and Kauper (2014), our control group spends about 60% of class time on lecture, 20% on instructor-led discussion, and 20% on other activities. Students in the treatment group study video lectures and complete an online quiz prior to class, work through scaffolded concept-based problem sets in groups with instructor support and complete an "exit" worksheet individually during class, and test their knowledge with a more challenging, though still low stakes, online quiz after class. We use assessment results from two principles of microeconomics classes taught by the same instructor during the spring 2017 semester. To best capture the effect of the RBIS, other aspects of the course, such as textbook, exams, and weights of course components, were consistent across the control and treatment groups. Students were unaware of any difference in the classes when registering. Our IRB approval allowed us to notify students and request their consent on the last day of class to prevent participation in the experiment from biasing their behavior.

The assessment data used in this study were collected at two distinct times: a pretest on the first day of class and the final exam. The final exam included embedded questions that were similar, but not identical, to those on the pretest. We collected data on five topics divided among three class content units: specialization and trade (Unit 1), demand and supply (Unit 1), elasticity (Unit 2), welfare analysis including taxes (Unit 2), and costs and competitive markets (Unit 3). All questions were open response, rather than multiple choice to make random guessing ineffective. The collection of assessment data at two points allows us to measure gains in learning that occur during the semester. Moreover, by separating the analysis into units, we can analyze the possibility that flipping will affect students differently either based on the nature of the material or based on the amount of time they have to adjust to the mode.

III. Descriptive Analysis

The data for this study are drawn from three sources: the course, university admissions profiles, and students' college transcripts. From these data, we construct two outcome variables: total learning gains (*LG*) and relative learning gains (*RLG*). Learning gains is an aggregate of indicator variables corresponding to whether a student incorrectly answered a question on the pretest but correctly answered it on the final exam. The *LG* variable is the sum of the learning gains for each question in the content unit. Similarly, to account for variations in the initial level of knowledge, we construct an alternate measure, relative learning gains (*RLG*), which is the difference of the percent of questions answered correctly on the final exam and on the pretest divided by the difference of 100% and the percent answered correctly on the pretest. *RLG* measures the learning gains of an individual as a percentage of the gains possible to the student. The key independent variable is the treatment status variable, which takes a value of one if the student was enrolled in the flipped RBIS class and zero otherwise. We control for prior coursework in economics and include background variables: gender, race/ethnicity, broad categories of college major, cumulative college credits earned, cumulative college GPA, citizenship status, whether the student resides on campus or if he/she commutes, SAT math score, and indicator variables that report whether the individual is from a low-income background and whether the student is a first generation college student.

A. Descriptive Analysis

The initial sample consists of 123 students, with 62 students enrolled in the control class and 61 enrolled in the treatment class. Because SAT math scores have been shown repeatedly to be an important predictor of success in economics classes (Olitsky and Cosgrove 2013, 2016; Cosgrove and Olitsky 2015), we eliminate students for whom no SAT scores are reported, reducing our sample from 123 to 106. There are students who enrolled in the class but dropped or withdrew from the class during the semester. Eliminating these observations from the sample, further reduces our sample size from 106 to 94.[1]

Comparing descriptive statistics across treatment status reveals several patterns. First, compared to students in the control group ($M = 5.14$, $SD = 2.93$), students in the treatment group answered significantly more of the repeated questions correctly on the Unit 1 topics ($M = 6.94$, $SD = 2.33$), answering 1.8 more questions correctly on average, $t(89) = 2.86$, $p = 0.005$. Students in the treatment group exhibited greater learning gains ($M = 6.44$, $SD = 2.22$) and relative learning gains ($M = 0.674$, $SD = 0.254$) than the control group (LG: $M = 4.92$, $SD = 2.93$; RLG: $M = 0.499$, $SD = 0.299$), $t(89) = 3.06$, $p = 0.001$. No significant differences in the learning gains for the Unit 2 and Unit 3 topics were observed. Total learning gains and relative learning gains across all topics were significantly higher for the treatment group (for LG, diff = 2.504; $p < 0.01$; for RLG, diff = 0.114; $p < 0.05$). Second, with three exceptions the means of the treatment and control groups were not statistically different across the measured controls. The treatment group had a significantly greater proportion of

[1] The students who did not take the final were equally distributed between the treatment and the control group, suggesting that students were not selecting out of the class because of its instruction type. These courses typically see a withdrawal/dropout rate of approximately 10%, so this sample reduction is reasonable.

students who were white (64% compared to 37%; diff = 0.277; $p < 0.01$), a significantly higher percentage of United States' citizens (98% compared to 78%; diff = 0.202; $p < 0.01$), and a significantly lower percentage of low income students (11% compared to 29%; diff = -0.175; $p < 0.05$).

IV. Econometric Specification

Because the treatment and control groups are statistically similar with regards to the observable controls, the effect of treatment can likely be captured by a standard OLS regression. Our benchmark econometric specifications are reduced forms of an educational production function, in which learning gains and relative learning gains are regressed on treatment status, and the aforementioned demographic and academic achievement controls. Separate regressions are estimated for each content unit, in addition to overall learning gains.

While the reduced form specification controls for characteristics that could potentially influence the effect of the treatment, it is possible that the OLS specification is not capturing the true effect of the treatment. Ideally, we would calculate the average treatment effect (ATE) by computing the average of the difference in learning gains between the treatment and control group. Yet, for each individual, we only observe their learning gains in their assigned treatment group, requiring us to estimate the counterfactual outcome, the learning gains a student in the treatment group *would have experienced* had he/she been assigned to the control group instead, or vice versa.

We employ Wooldridge's "Double Robust" estimator (Wooldridge 2007, 2010 p 930-934), which combines two methods of determining the counterfactual: regression adjustment and propensity score weighting. Combining these two methods makes the resulting estimates more robust to specification error, a key econometric concern surrounding program evaluation methods.

The double robust estimator is implemented as follows. First, we estimate the propensity score, the probability that an individual is in the treatment group, using a probit regression, with the controls being all of the independent variables included in the benchmark OLS specification.[2] Second, we

[2] Caliendo and Kopeinig (2008) presents a detailed discussion of the assumptions required to estimate and use a propensity score to which the present analysis conforms. We test these assumptions as prescribed in the literature; however due to space restrictions we do not report the results of these tests in this article. These are results are available upon request.

estimate separate weighted linear regressions for the treatment and control groups of the learning gains outcome on the relevant covariates, using the inverse of the propensity score as the regression weight for the treatment group, and the inverse of one minus the propensity score as the regression weight for the control group. Third, the *ATE* is estimated by computing the average of the difference in the linear predictions from the weighted regressions.

Because we use both propensity score weighting and regression adjustment, the estimates are robust to specification error either of the regression model or of the propensity score estimation. Wooldridge (2007, 2010) shows that only one of the two must be specified correctly for the estimation to provide consistent estimates of the *ATE*. The estimation procedure also estimates the average treatment effect on the treated (*ATT*), computing the average treatment effect for only those in the treatment group and the potential outcome means, which are the linear predictions of the average learning gains and average relative learning gains.

V. Results

The OLS results suggest that participation in treatment is significantly associated with learning gains, $b = 2.504$, $t(79) = 3.19$, $p < 0.01$, and relative learning gains, $b = 0.105$, $t(79) = 2.53$, $p < 0.05$. Examining the learning gains by content unit, the results indicate that for Unit 1 questions, participation in the RBIS treatment is significantly associated with additional improvements both in learning gains, $b = 1.526$, $t(79) = 2.88$, $p < 0.01$, and in relative learning gains, $b = 0.168$, $t(79) = 3.02$, $p < 0.01$, compared to the control group, on average. Likewise, for Unit 2, participation in the treatment is significantly associated both with additional learning gains, $b = 1.021$, $t(79) = 2.30$, $p < 0.05$ and with additional relative learning gains $b = 0.113$, $t = 2.30$, $p < 0.05$. For Unit 3, there are significant additional gains associated with treatment neither with learning gains, nor with relative learning gains.

The OLS results also reveal two patterns that are consistent with previous studies and our prior expectations. First, there is a significant, positive association between learning gains and college GPA, of roughly the same magnitude of the treatment effect for the overall learning gains, $b = 2.79$, $t(79) = 4.50$, $p < 0.001$), and relative learning gains $b = 0.125$, $t(79) = 3.92$, $p < 0.001$, and for Units 1 (*LG*: $b = 1.298$, $t(79) = 3.25$, $p < 0.001$: *RLG*: $b = 0.133$, $t(79) = 3.11$, $p < 0.001$) and 2 (*LG*: $b = 1.49$, $t(79) = 4.25$, $p < 0.001$: *RLG*: $b = 0.166$, $t(79) = 4.25$, $p < 0.001$). College GPA has no significant association with learning gains for

Unit 3 (*LG*: b = 0.393, t(79) = 1.42, $p > 0.1$: *RLG*: b = 0.054, t(79) = 1.32, $p > 0.1$). Second, there is a significant, positive association between SAT math scores and learning gains (*LG*: b = 0.022, t(79) = 3.56, $p < 0.001$: *RLG*: b = 0.001, t(79) =3.78, $p < 0.001$). Overall, increasing SAT math scores by one standard deviation (approximately 81 points) is associated with an increase in learning gains of 1.78 questions or approximately 9.5%.

The ancillary propensity score estimation has some noteworthy results. Of the controls in the propensity score estimation, two are statistically significant. First, students who are US citizens are also more likely to be in the treatment group, b = 2.53, t(79) = 2.29, $p < 0.05$. Second, students in the treatment group are less likely to be from a low-income family, b = -1.528, t(79) = 2.002, $p < 0.05$. The remaining coefficients are insignificant.

The estimates from the double robust estimation procedure confirm and strengthen the results from the OLS model, indicating a significant, positive effect of treatment on overall learning gains (*LG*: b = 3.27, z = 4.31, $p < 0.001$: *RLG*: b = 0.149, z =3.97, $p < 0.001$) and learning gains in the first content unit (*LG*: b = 1.898, z = 3.83, $p < 0.001$: *RLG*: b = 0.211, z =4.21, $p < 0.001$) and the second content unit (*LG*: b = 1.374, z = 3.45, $p < 0.01$: *RLG*: b = 0.153, z =3.45, $p < 0.001$), and no significant treatment effect for the Unit 3 content (*LG*: b = 0.367, z = 1.06, $p > 0.1$: *RLG*: b = 0.069, z = 1.37, $p > 0.1$). These results indicate that the average effect of treatment is an additional 3.2 questions or additional relative learning gains of approximately 15%. The potential outcome means results indicate that students in the treatment group answered 11.6 more questions correctly or 62.6%, on average, on the final exam than they did on the pretest; whereas the control group improved by 8.4 questions or 47.7%, on average.

V. Discussion

This study tests the effectiveness of instructional strategies espoused by the cognitive science literature, comparing the average learning gains between students in a “traditional” principles of microeconomics course to the average learning gains of students in a course designed to deliberately employ these strategies. Two econometric specifications generate similar results: significant positive treatment effects for the overall learning gains, and for the first and second course content units.

These results are consistent with the cognitive science literature. Repeated testing of a concept, requiring students to recall concepts at spaced intervals, is one effective learning strategy. It is possible that students in the RBIS-based treatment group learned more than students in the control group because the content in the first two units was tested more than the last unit.

The results of this study suggest two key recommendations for instructors looking to improve student outcomes in economics. First, implementing an RBIS-based course is a promising teaching strategy and one that likely would lead to improved student outcomes. Second, "flipping" makes it easier for these strategies to be employed. Allowing the instructor to spend more time interacting with students while they are working on difficult material conforms to several of the "best practice" strategies. For example, when the instructor moves around the classroom answering students' questions, she is providing specific formative feedback, helping students to build a mental model, directly confronting students' challenges with terminology, having students explain why they chose their answers and having students describe their thought process (metacognition). While flipping is not a prerequisite to employ these strategies, it certainly makes it easier to do so.

REFERENCES

Brown, Peter C., Henry L Roediger III, and Mark A. McDaniel. 2014. *Make It Stick: The Science of Successful Learning*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Caliendo, Marco, and Sabine Kopeinig. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *Journal of Economic Surveys* 22 (1):31–72.

Cosgrove, Sarah B., and Neal H. Olitsky. 2015. "Knowledge Retention, Student Learning, and Blended Course Work: Evidence from Principles of Economics Courses: Knowledge Retention and Mode of Study." *Southern Economic Journal* 82 (2):556–79.

Goffe, William L., and David Kauper. 2014. "A Survey of Principles Instructors: Why Lecture Prevails." *The Journal of Economic Education* 45 (4):360–75.

Olitsky, Neal H., and Sarah B. Cosgrove. 2013. "The Effect of Blended Courses on Student Learning: Evidence from Introductory Economics Courses." *International Review of Economics Education*, November.

———. 2016. "The Better Blend? Flipping the Principles of Microeconomics Classroom." *International Review of Economics Education* 21 (January):1–11.

Wooldridge, Jeffrey M. 2007. "Inverse Probability Weighted Estimation for General Missing Data Problems." *Journal of Econometrics* 141 (2):1281–1301.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, Mass: MIT Press.