Measuring the Effect of Blended Learning: Evidence from a Selective Liberal Arts College

By Aaron Swoboda and Lauren Feiler*

We increased the amount of time spent in our introductory microeconomics courses on active-learning pedagogies like group-problem solving and simulations. In order to make time for these activities, we employed a blended learning approach that shifted much of the basic content delivery of material outside of the class. In this paper we explore some of the effects of this shift, by asking whether the potential increase in higher-level learning outcomes brought about by the class activities were offset by reduced basic content knowledge without the in-class content delivery.

To answer this question we employed a difference-in-differences approach comparing the pre- and posttest scores on the Test of Understanding in College Economics (TUCE; Walstad, Watts and Rebeck, 2007) across students in the blended and control sections. Students in the four blended sections of the course increased their TUCE scores by *more* than the students in the four control sections even after controlling for basic student demographic characteristics. This difference also persists after controlling for students' self-reported measures of attitude regarding the examination process via the Student Opinion Survey (SOS; Sundre, 2007).

I. Related Studies

Studies of online learning in the economics literature are often focused on the benefits of the use of technology as an enhancement to traditional courses (e.g. Agarwal and Day, 1998; Hernandez-Julian and Peters, 2012; Lee, Courtney and Balassi, 2010; Sosin et al., 2004), typically finding small positive effects of technology use on student performance. While these studies point to potential benefits of online components to a course, few directly compare blended courses in economics to more traditional courses that may also include online components.

Brown and Liedholm (2002) compare face-to-face, hybrid, and purely online settings for undergraduate microeconomics courses, while Terry and Lewer (2003) make the same comparison for graduate students in macroeconomic theory and international economics courses. Neither study shows a significant difference between performance on common final exam questions in the face-to-face and hybrid courses. Both find lower performance in the online course compared to the face-to-face course, and this pairwise comparison is also significant between the online and hybrid course in Terry and Lewer. Olitsky and Cosgrove (2014) focus solely on blended versus traditional courses, and they account for potential selection bias into a blended course using propensity score matching. They support the earlier findings of no differences between outcomes in blended and face-to-face courses.

Our study differs from the above in our definition of blended learning. While these studies use online components to replace a portion of the face-to-face time spent in the classroom, our approach was to maintain the same class schedule while altering what was covered in the classroom. Our courses are taught in ten-week terms, so we did not wish to further limit face-to-face contact with students. Much of what we do would be considered "flipping" the classroom. Roach (2014) shows that students perceive this technique to be effective in an

* Swoboda: Carleton College, 1 N. College St., Northfield, MN 55057, aswoboda@carleton.edu. Feiler: Carleton College, 1 N. College St., Northfield, MN 55057, lfeiler@carleton.edu. This work was funded by a Future Learning Technologies grant and Curricular Innovation Grant at Carleton College. We thank Andrea Nixon, Sara Nielsen, and the other faculty in our department.

introductory microeconomics course; we set out to quantify whether it is.

II. Methodology

We studied eight sections of Principles of Microeconomics at Carleton College during the 2014-15 academic year. Each section contained between 20 and 30 students for a total of roughly 200 (just over 10 percent of the Carleton student body). Four sections were taught by the authors with blended learning techniques, while four control sections were taught by three other professors with more traditional styles. The blended sections required students to read textbook chapters, watch videos of lecture material created by the authors, and answer basic comprehension questions online before coming to a class session on a topic. During the class session, the professors began by answering questions about the out-of-class materials and giving mini-lectures targetted to troublesome items, but students spent much of their time engaged in group problem-solving, simulations and discussion activities. After the session, students were individually assigned a second round online questions on that day's topic and preparatory materials for the next topic. Online homework assignments were completed using Sapling Learning and provided instant grading and feedback. The control courses primarily used "chalk-and-talk" and traditional written homework assignments.

To compare outcomes, we administered the 4th edition of the Microeconomics TUCE in the first and last weeks of the ten-week term. Students took the test outside of class and were provided with a small incentive to participate, such as the ability to drop the lowest homework grade. We described the TUCE as a survey rather than a test, and students knew that their scores were recorded anonymously so professors could not link performance to a course grade. Participants also filled out the Student Opinion Survey after completing the TUCE. The SOS uses ten questions to assess the level of effort students put into the survey they just took, as well as how important they perceive the survey to be. All surveys were provided in sealed envelopes and marked with an ID number. Our Director of Educational Research used a correspondence—unknown to us—between IDs and names to link the survey results and students' demographic and educational data and returned an anonymized dataset to us. The specific demographic variables included standardized test scores, gender, expected college graduation year, financial need and previous experience with economics (either in High School or at the College level).

We found no significant differences in student demographics across the treatment and control sections of the course. We also found the demographic characteristics for those students who elected to participate in our study were little different from the students who chose not to take the surveys; the only statistically significant difference was that females were more likely to participate than males. We restricted our dataset to only those 150 students that took the TUCE at the beginning and end of the term.

III. Data and Results

Table 1 contains information on students' TUCE performance, survey responses, and demographic characteristics. The mean student scored 20.5 points on the TUCE at the end of the term compared to 15.0 on the pretest. The difference in TUCE scores is significantly different from zero at conventional levels. It should be noted that the average pretest TUCE score of 15 points is above the national average of 12.8 on the *posttest*. Although students, on average, performed better on the TUCE at the end of the term compared to the beginning, they surprisingly did not report putting in more effort or attributing more weight to that survey at the end of the term. On the SOS, students reported roughly neutral "importance" levels (approximately 15.5 on a scale ranging from 5 to 25) and moderately high effort (approximately 18.3 on the same scale) on both the pretest and posttest. We also report summary statis-

Table 1—Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
TUCE Score					
Pretest	150	15.1	4.4	4	28
Posttest	150	20.7	4.4	10	29
Difference	150	5.5	3.8	−4	18
SOS Importance Subscale					
Pretest	149	15.3	3.9	5	25
Posttest	150	15.7	3.8	5	25
Difference	149	0.4	3.5	−14	14
SOS Effort Subscale					
Pretest	149	18.5	3.4	5	25
Posttest	150	18.1	3.3	5	25
Difference	149	−0.3	3.0	−7	13
Demographic Characteristics					
Standardized Test Percentile	150	93	7	57	99
SAT Math	104	719	64	500	800
SAT Verbal	104	694	69	400	800
ACT Composite	72	31	2.7	23	36
HS Macro (Yes = 1)	150	0.1	0.3	0	1
HS Micro (Yes = 1)	150	0.1	0.3	0	1
College Macro (Yes = 1)	150	0.4	0.5	0	1
Expected Graduation Year (20XX)	150	17.4	0.9	15	18
Gender (Male = 1)	150	0.5	0.5	0	1
Minority or Int'l Student (Yes = 1)	150	0.4	0.5	0	1

tics of participants' demographic characteristics. The typical student would be a first-year student with standardized test scores in roughly the 93rd percentile and no prior experience with economics.

Table 2 shows regression specifications modeling the change in TUCE scores. In all cases, we see a positive and statistically significant coefficient on the blended treatment variable. Columns (1) to (3) use the difference in the pre- and posttest TUCE as the dependent variable. Column (1) shows that students in the traditional courses improved by roughly 4 points out of 30, while students in the blended courses improved by 6 points. Column (2) adds the importance and effort subscales of the SOS. Though importance is marginally significant, the magnitude is small in relation to the observed changes in how important students deemed the posttest to be in comparison to the pretest. An increase of 1 point on the TUCE corresponds to a 6 point increase in the importance rating, and only 7 students in our sample showed that rating increase. The specification represented in Column (3) also included the demographic variables described above. These did not affect the main results and coefficients are not included here due to space constraints. (See Data Appendix for full results.)

Columns (4) to (6) use a gap closing measure as the dependent variable to control for students' pretest score. This measure is the difference between the TUCE pre- and posttest over the difference between the pretest and the maximum possible score of 30. In columns (1) to (3), two students who improved by 5 points would look the same regardless of pretest score, whereas the initial score matters with the gap closing measure. A student who improved from 10 points to 15 points would close 25% of the gap between her score and the maximum, while a student who improved from 20 to 25 points would close 50% of the gap between the initial score and the 30 points possible. Our specification in column (4) shows that students on average reduced the gap by 31% in the traditional courses and an extra 9% in the blended courses. Columns (5) and (6) parallel (2) and (3) and include the SOS variables and the additional demographic variables. Across these regressions,

Table 2—Regression Results

	Dependent variable:					
	Δ TUCE = Posttest - Pretest			Δ TUCE / (30 - Pretest)		
	(1)	(2)	(3)	(4)	(5)	(6)
Blended Treatment: Yes	1.92***	1.74***	1.82***	0.09**	0.08*	0.08*
	(0.62)	(0.61)	(0.64)	(0.04)	(0.04)	(0.04)
Difference in Effort		0.06	0.09		0.003	0.01
		(0.10)	(0.10)		(0.01)	(0.01)
Difference in Importance		0.17*	0.12		0.01	0.01
		(0.09)	(0.09)		(0.01)	(0.01)
Constant	4.34***	4.34***	11.38	0.31***	0.31***	0.75
	(0.49)	(0.48)	(7.92)	(0.03)	(0.03)	(0.54)
Demographic Controls Included	No	No	Yes	No	No	Yes
Observations	150	149	149	150	149	149
R^2	0.06	0.09	0.18	0.03	0.05	0.13
Adjusted R^2	0.05	0.07	0.09	0.02	0.03	0.04

Note: Demographic control coefficients are not shown here. The controls were: experience in previous college or high school economics courses, standardized test percentile, expected graduation year, gender, minority or international student status, and financial need.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

we see that including student self-reported effort and weight placed on the test had no effect on TUCE improvement.

IV. Conclusion

The combination of online homework, video-based lecture material, and in-class problem solving that we used resulted in higher TUCE scores for the blended courses than for the controls. This stands in contrast to the few studies of blended courses within economics, so we consider some potential explanations here. One possibility is that we maintained the same amount of class time in blended and non-blended courses, so students do not give up face-to-face interaction with increased online work. Another is that our participants are drawn from an unusual student population, averaging within the top 10 percent of high-schoolers on standardized tests and starting the course with a fair amount of economic intuition, as measured by the TUCE. Another possibility is that we were biased as researchers and "taught to the test." This could be a concern since we were unable to conduct our study with the same instructors teaching both traditional and blended courses, but the anonymity of the study helped ensure that we could not spend extra time with students who scored low on the pretest, and a shared set of coverage goals across the department that need to be met within the 10-week term would make it difficult to cover TUCE questions that are not part of the usual curriculum.

Other studies (Brown and Liedholm, 2002; Emerson and Taylor, 2004, and Dickie, 2006 on the use of experiments) find evidence that certain student subgroups, particularly higher-achieving students, benefit more from non-traditional techniques. Preliminary analysis does not show interaction effects between our treatment and high standardized test scores, but given the selective nature of our institution, we may not have the variation necessary to satisfactorily test the hypothesis.

Future researchers may benefit from the finding that students' self-reported effort on the TUCE (as measured by SOS scores)

does not vary between the pretest and the posttest, and they attribute similar importance to both tests despite the posttest being administered soon before a final. One might worry that part of the typical improvement in TUCE scores is simply that students care more about the test once they have actually taken the course, but we find that this is not the case. Although a student putting more value in the TUCE is linked to higher TUCE scores in one of our regression specifications, this effect is very small, and would have little impact anyway since students on the whole do not place more weight on the posttest. In future work, we plan to explore the question of potential interaction effects more, and whether a grade incentive would affect students' answers on the TUCE, using a within-subjects methodology based on a subset of TUCE questions that we included on final exams.

REFERENCES

Agarwal, Rajshree, and A. Edward Day. 1998. "The Impact of the Internet on Economic Education." *The Journal of Economic Education*, 29(2): 99.

Brown, Byron W., and Carl E. Liedholm. 2002. "Can Web Courses Replace the Classroom in Principles of Microeconomics?" *American Economic Review*, 444–448.

Dickie, Mark. 2006. "Do Classroom Experiments Increase Learning in Introductory Microeconomics?" *The Journal of Economic Education*, 37(3): 267–288.

Emerson, Tisha L. N., and Beck A. Taylor. 2004. "Comparing Student Achievement across Experimental and Lecture-Oriented Sections of a Principles of Microeconomics Course." *Southern Economic Journal*, 70(3): 672–693.

Hernandez-Julian, Rey, and Christina Peters. 2012. "Does the Medium Matter? Online versus Paper Coursework." *Southern Economic Journal*, 78(4): 1333–1345.

Lee, William, Richard H. Courtney, and Steven J. Balassi. 2010. "Do Online Homework Tools Improve Student Results in Principles of Microeconomics Courses?" *American Economic Review*, 100(2): 283–286.

Olitsky, Neal H., and Sarah B. Cosgrove. 2014. "The Effect of Blended Courses on Student Learning: Evidence from Introductory Economics Courses." *International Review of Economics Education*, 15: 17–31.

Roach, Travis. 2014. "Student Perceptions Toward Flipped Learning: New Methods to Increase Interaction and Active Learning in Economics." *International Review of Economics Education*, 17: 74–84.

Sosin, Kim, Betty J. Blecha, Rajshree Agarwal, Robin L. Bartlett, and Joseph I. Daniel. 2004. "Efficiency in the Use of Technology in Economic Education: Some Preliminary Results." *American Economic Review*, 253–258.

Sundre, Donna L. 2007. "The Student Opinion Scale (SOS): A Measure of Examinee Motivation." *James Madison University, The Center for Assessment & Research Studies.*

Terry, Neil, and Joshua Lewer. 2003. "Campus, Online, or Hybrid: An Assessment of Instruction Modes." *Journal of Economics and Economic Education Research*, 4(1): 23–34.

Walstad, William B., Michael Watts, and Ken Rebeck. 2007. "Test of Understanding in College Economics, Fourth Edition." *National Council on Economic Education.*

Data Appendix

A1. Complete Regression Results

Due to space constraints, we did not include all of the regression covariates in Table 2. We show the full table of results in Table A1.

A2. Additional Variable Descriptions

Carleton College allows students to report either SAT or ACT scores in their admission file. For comparability, we created a Standardized Test Percentile variable by converting test scores into percentiles and taking the mean percentile across the different reported standardized tests for each student. Percentile data was obtained from `https://secure-media.collegeboard.org/digitalServices/pdf/sat/sat-percentile-ranks-crit-reading-math-writing-2014.pdf` and `http://www.actstudent.org/scores/norms1.html`.

We also collected a financial need variable from our Admissions Office. Each student is designated as one of five need categories. The highest need group was determined to have financial need greater than three-quarters of the comprehensive fee at Carleton College ($60,102 for the 2014-15 academic year). The second highest need group had need greater than half of the comprehensive fee. The third (fourth) highest need categories had need greater than one fourth (zero) of the comprehensive fee. The final group was determined to have no financial need or did not apply for aid. Financial need is bimodally distributed. Approximately half of students received no financial aid, while one quarter of the students were in the highest need category (the remaining quarter of students is distributed across the remaining three aid categories).

A3. Covariates by Subgroups

Because our study is observational, we tested for differences in student characteristics across the control and treatment class sections (Tables A2 and A3) as well as those students who opted out of the study vs those that participated (Tables A4 and A5). No statistically significant differences were found across these groups with the exception of participation by gender. Males were significantly more likely to opt out of the study than females.

Table A1—Full Regression Results

	Dependent variable:					
	Δ TUCE= Posttest - Pretest			Δ TUCE/(30 - Pretest)		
	(1)	(2)	(3)	(4)	(5)	(6)
Blended Treatment: Yes	1.92***	1.74***	1.82***	0.09**	0.08*	0.08*
	(0.62)	(0.61)	(0.64)	(0.04)	(0.04)	(0.04)
Difference in Effort		0.06	0.09		0.003	0.01
		(0.10)	(0.10)		(0.01)	(0.01)
Difference in Importance		0.17*	0.12		0.01	0.01
		(0.09)	(0.09)		(0.01)	(0.01)
Completed College Macro			−0.12			−0.01
			(0.69)			(0.05)
Completed HS Macro			−0.10			0.05
			(1.04)			(0.07)
Completed HS Micro			−0.59			−0.003
			(1.33)			(0.09)
Stand. Test Percentile			−0.07			0.0001
			(0.05)			(0.003)
Year of College			0.05			−0.02
			(0.38)			(0.03)
Male: Yes			−0.69			−0.04
			(0.63)			(0.04)
Minority or Int'l Student			−1.10			−0.08*
			(0.70)			(0.05)
Financial Need: Low			0.30			−0.01
			(1.37)			(0.09)
Financial Need: Some			−1.69			−0.11
			(1.07)			(0.07)
Financial Need: More			−2.31**			−0.15**
			(1.02)			(0.07)
Financial Need: Most			−0.84			−0.04
			(0.78)			(0.05)
Constant	4.34***	4.34***	11.38	0.31***	0.31***	0.75
	(0.49)	(0.48)	(7.92)	(0.03)	(0.03)	(0.54)
Observations	150	149	149	150	149	149
R^2	0.06	0.09	0.18	0.03	0.05	0.13
Adjusted R^2	0.05	0.07	0.09	0.02	0.03	0.04

Table A2—Demographic Statistics by Class Sections

variable	group	n	mean	sd	median	min	max
Standardized Test Percentile	Control	91	94.1	6.6	95.5	57	99
	Treatment	102	93.1	7.3	95.8	56	99
SAT Math	Control	54	727.4	58.0	740	600	800
	Treatment	71	713.2	65.5	720	500	800
SAT Verbal	Control	54	696.1	70.8	695	400	800
	Treatment	71	694.2	67.5	700	500	800
ACT Composite	Control	49	31.6	2.6	32	23	36
	Treatment	52	31	2.8	31.5	21	35
HS Macro (Yes = 1)	Control	91	0.1	0.4	0	0	1
	Treatment	102	0.1	0.3	0	0	1
HS Micro (Yes = 1)	Control	91	0.1	0.3	0	0	1
	Treatment	102	0.1	0.2	0	0	1
College Macro (Yes = 1)	Control	91	0.4	0.5	0	0	1
	Treatment	102	0.4	0.5	0	0	1
Exp. College Grad Year	Control	91	17.5	0.8	18	15	18
	Treatment	102	17.4	0.9	18	15	19
Gender (Male = 1)	Control	91	0.6	0.5	1	0	1
	Treatment	102	0.6	0.5	1	0	1
Minority or Intl Student	Control	91	0.4	0.5	0	0	1
	Treatment	102	0.3	0.5	0	0	1

Table A3—Frequency of Student Financial Need by Class Section

	No Need	Lowest Need	Some Need	More Need	Most Need
Control Sections	42	7	6	13	23
Treatment Sections	51	2	11	11	27

Table A4—Student Demographic Statistics by Participation Decision

variable	group	n	mean	sd	median	min	max
Standardized Test Percentile	In Study	150	93.3	6.8	95.5	57	99
	Opted Out	43	94.4	7.4	96.5	56	99
SAT Math	In Study	104	718.5	64.3	740	500	800
	Opted Out	21	723.8	54.2	720	600	800
SAT Verbal	In Study	104	694.7	69.2	705	400	800
	Opted Out	21	696.7	67.2	690	570	800
ACT Composite	In Study	72	31.1	2.7	31	23	36
	Opted Out	29	31.8	2.7	32	21	35
HS Macro (Yes = 1)	In Study	150	0.1	0.3	0	0	1
	Opted Out	43	0.1	0.4	0	0	1
HS Micro (Yes = 1)	In Study	150	0.1	0.3	0	0	1
	Opted Out	43	0.1	0.3	0	0	1
College Macro (Yes = 1)	In Study	150	0.4	0.5	0	0	1
	Opted Out	43	0.5	0.5	1	0	1
Exp. College Grad Year	In Study	150	17.4	0.9	18	15	18
	Opted Out	43	17.4	0.9	18	15	19
Gender (Male = 1)	In Study	150	0.5	0.5	1	0	1
	Opted Out	43	0.8	0.4	1	0	1
Minority or Intl Student	In Study	150	0.4	0.5	0	0	1
	Opted Out	43	0.3	0.4	0	0	1

Table A5—Frequency of Student Financial Need by Study Participation Decision

	No Need	Low Need	Some Need	More Need	Most Need
In Study	73	8	14	17	38
Opted Out	20	1	3	7	12