# The Effect of Blended Courses on Student Learning: Evidence From Introductory Economics Courses

Neal H. Olitsky[1] Sarah B. Cosgrove[2]

University of Massachusetts Dartmouth

November 19, 2012

**ABSTRACT**

Over the past decade there has been a large increase in the number of colleges and universities that offer fully online courses and blended courses (courses with a face-to-face component and with an online component). The number of students enrolling in these courses has also increased. These courses are less costly for universities to offer and provide students with more flexibility than traditional classes. However, the effect of these courses on student learning remains unclear. This study examines the effect of blended learning on a specific student learning outcome in introductory economics courses. The effect of blending on learning is determined by comparing scores on quizzes and exams between students in a blended course (the treatment) and students in a traditional face-to-face course (the control). This study accounts for the potential bias due to non-random selection into treatment by using propensity score matching. The results indicate no significant effects of blending on student learning.

**JEL Classification**: A22; I20; I21

**Keywords**: Educational Economics, Propensity Score Matching, Teaching of Economics, Blended Learning

Acknowledgements: This research was funded by a Davis Educational Foundation grant. The authors thank Jeannette Riley and Catherine Gardner, co-PIs of the grant, for their assistance throughout the project.

**DO NOT CITE WITHOUT PERMISSION**

[1] E-mail: neal.olitsky@umassd.edu; University of Massachusetts Dartmouth, Department of Economics, 285 Old Westport Rd, North Dartmouth MA, 02747

[2] E-mail: sarah.cosgrove@umassd.edu; University of Massachusetts Dartmouth, Department of Economics, 285 Old Westport Rd, North Dartmouth MA, 02747

There has been a shift in higher education toward more fully online and blended courses in recent years. While they have been defined differently in the literature (Williams 2002, Garrison, Kanuka, and Hawes 2002, Garnham and Kaleta 2002), it is commonly accepted that blended or hybrid courses integrate traditional face-to-face class sessions with online class components which take the place of some class time. An extensive Sloan Consortium survey found that in 2004 almost 55 percent of institutions offered at least one blended course. Moreover, in the same year 79 percent of public institutions offered at least one blended course (Allen, Seaman, and Garrett 2007).

Given this extensive shift toward online and blended learning, the question of primary importance is the level of student learning in this setting compared to a traditional face-to-face (F2F) course. While the research on this question is extensive across disciplines, it is very limited with respect to economics. This study draws from the advisory literature on how best to develop and deliver a blended course and from other disciplines on assessment of learning in a blended course to fill a gap regarding the effectiveness of blended learning in economics.

To determine the effectiveness of blended coursework, we compare the learning outcomes for students enrolled in principles of economics courses (both micro and macro) in the 2011/2012 academic year. Of the seven sections that were examined, two were blended. The remaining sections serve as the control group, and were F2F courses. Data on student performance were matched with university transcript and enrollment information to provide additional controls. Using these data, we estimated the differences in learning outcomes across modes to determine the effect of blending on learning.

This study will advance the literature in the field in four ways. First, it is a current study of blended learning in economics, informed by blended learning literature across disciplines. Second, our methodology controls for the selection bias found in previous studies of online versus F2F learning, but for which previous studies on blended learning fail to account. We use propensity score matching (PSM) to recover the causal effect of blending on student learning. Further, we provide evidence that the PSM specification accounts for the non-random selection into treatment by estimating Imbens bounds (Imbens 2003) to determine how important an unmeasured selection variable must be to undermine our conclusions. Third, we clearly specify the similarities and differences between the blended courses and the F2F courses. Finally, we target and assess a specific student learning objective.

With few exceptions, our findings suggest no significant difference in learning between blended and F2F sections. Most of the estimated treatment effects are both statistically insignificant and small in magnitude, regardless of estimation technique.  While these results are consistent with the literature, our results also suggest the presence of sizeable selection bias; in most cases, OLS estimates of the treatment effect overstate the effect of blending.

The paper proceeds as follows: section two reviews the literature, section three outlines the experiment and data collection, section four explains the data and descriptive statistics, section five discusses the estimation strategy, section six provides the results, and section seven concludes.

## LITERATURE REVIEW

The vast literature on the effects of blended course delivery reveals a mix of benefits and costs.  On the positive side, evidence shows that blended courses provide students with time flexibility and improved learning outcomes, afford more student-teacher interaction, increase student engagement, allow for continuous improvement in a course, enhance an institution's reputation, expand access to educational offerings, and reduce operating costs (Vaughan 2007).  Consistent with Vaughan's findings, a meta-analysis across disciplines conducted by the US Department of Education found statistically stronger learning outcomes in blended classes compared to F2F classes (Means et al. 2009).  Moreover, Arbaugh et al. (2009) conducted a meta-analysis specific to the use of online and blended learning in the business disciplines.  They found that as online and blended learning courses become more prevalent, any negative performance differences between F2F and online or blended classes diminished or shifted to favoring the latter, suggesting a learning curve in development and completion of technology-assisted courses. However, the benefits of blending are not without costs.  Some of the documented costs are students' struggles with time management and responsibility for their own learning, difficulty using new technology, an increased time commitment from faculty, inadequate professional development support, resistance to organizational change, and difficulties establishing a supportive culture for blended learning (Vaughan 2007).

Compared to other disciplines and the literature as a whole, there are relatively few studies on blended or fully online delivery in economics.  Of these studies, there are two primary categories: web-based enhancements to F2F classes and comparisons of the outcomes in online or blended versus F2F classes.  The first category is of limited interest to this report and includes studies that provide examples of the use of technology to enhance a F2F class but not to provide a blended experience.  The second group of studies compares the results from online and/or blended courses with F2F courses.  Since 2000, all but one of the studies that compare fully online with F2F courses in economics found that students learn less in fully online courses.  These results persist in undergraduate (Brown and Liedholm 2002; Coates et al. 2004) and graduate courses (Anstine and Skidmore 2005; Terry and Lewer 2003). Only Navarro and Shoemaker (2000) found improved learning performance in the form of higher final exam scores in the online sections compared to F2F sections. Notably, Coates et al. (2004) report substantial selection bias in the students who choose to enroll in fully online courses. Failure to account for this selection bias would have resulted in a conclusion that there was no significant difference in learning from the two different modes of delivery.

Only two studies were found that formally and quantitatively assess the efficacy of blended learning in economics. Both of these studies compare three modes: fully online, blended, and F2F, but only find significant differences between the fully online and F2F results. Brown and Liedholm (2002) studied undergraduate principles of microeconomics classes while Terry and Lewer (2003) studied graduate students in macroeconomic theory or international economics courses. Both studies concluded that there was no statistically significant difference between student performance on the final exam in the F2F courses and the blended courses. However, neither study controlled for selection bias by mode. Given the limited and somewhat dated research on the effects of blended learning in economics, a current study that controls for selection bias is warranted.

## EXPERIMENT AND DATA COLLECTION

The authors became involved in blended learning through a grant from the Davis Educational Foundation to support a campus-wide teaching initiative called the Implementation of Blended Learning for the Improvement of Student Learning (IBIS). Participation in this program involved completion of a faculty development course in summer 2011 that taught best practices in blended learning and required development of blended courses to be taught in the fall 2011 semester. Labeling the instructors 1 and 2, in fall 2011, Instructor 1 taught one blended section of principles of microeconomics, and Instructor 2 taught two sections of principles of macroeconomics, one blended and one F2F. In spring 2012, both instructors taught two F2F sections of principles (Instructor 1 taught micro and Instructor 2 taught macro) Student data and performance from these courses are used in this study.

For the purposes of this study, we adopt the definition of blending discussed in Allen, Seaman, and Garrett (2007), a report provided by the Sloan Consortium. According to this report, a blended course is one that delivers between 30% and 79% of its content online and "typically uses online discussions, and typically has some face-to-face meetings" (Allen, Seaman, and Garrett 2007, p.5) For the blended sections, both instructors used a 2/3-1/3 blend; each blended course substituted online instruction for approximately one third of the semester's class periods.

The online instruction included online lectures, article analyses, discussion board assignments and group wiki assignments. Given the amount both of online instruction and of online coursework, the blended courses in this study falls within the definition described above.

Due to the commonalities in the first several weeks of the principles of macro and micro courses, we designed our blended and F2F courses the same way and are able to assess our results both within and across courses. To maximize consistency across sections, we used the same textbook, online homework management website, course management system, assignments, and exam. For the purpose of this study, we chose one key learning outcome that was specific to the first unit but central to all of economics: Compute and compare opportunity costs of different decision-makers to determine the most efficient specialization of production.

For the target SLO, all students were required to complete 1.) a chapter reading from the textbook, 2.) a pair-and-share practice problem set (F2F), and 3.) an assignment from an online homework management website with interactive graphs, tables, and corresponding questions (online). The students in the blended sections had two fewer 75 minute F2F class during this unit. In place of the two classes, students were assigned three online exercises. First, students were required to read an article applying the opportunity cost concept and complete follow-up discussion board questions online. Second, students were required to complete a small group wiki project for which students were required to develop their own comparative advantage example with computations and analysis and post it online for their classmates to review. Finally, students were required to participate in an online market experiment. Copies of these assignments are available upon request from the authors. Students in all courses were given an identical exam covering the unit. The exam questions were a combination of multiple choice and free response questions.[3]

Each student's assessment results were matched to his/her college transcript data and demographic information, allowing us to control both for academic achievement and for each student's background. The transcript data reported each course in which a student enrolled and the grades they earned for each course. The demographic data

[3] For each question, we compute the difficulty index, the proportion of students who correctly answered the question, and the discrimination index, the difference between the percentage of the upper group of students who correctly answered the question and the percentage of the lower group of students who correctly answered the question. Consistent with Kelley (1939) the upper group is the highest scoring 27% and the lower group is the lowest scoring 27% of students. For the SLO-specific multiple choice questions, the difficulty index values ranged from 0.31 to 0.68 with a mean value of 0.46. The discrimination index values for the same questions ranges from 0.03 to 0.52 with a mean value of 0.17. If we treat each part of the SLO-specific free response question as a separate entity and count as correct only responses that received full credit, the difficulty index values of these questions range from 0.07 to 0.91 (mean 0.66) and the discrimination index values range from 0.15 to 0.56 (mean 0.34).

included: program of study, college major, cumulative credits earned, cumulative GPA, US citizenship status, first term enrolled at the university, SAT Math and Verbal scores, and race/ethnicity. The initial sample consisted of 354 observations. However, because we wanted to control for academic achievement, we eliminated 36 observations for which no SAT math or verbal scores were reported. The final sample consisted of 318 observations.

## DATA AND DESCRIPTIVE ANALYSIS

We examine first the differences in means between the blended observations and F2F observations for all of the variables of interest. Table 1 reports both the descriptive statistics and the differences in means between blended and F2F sections. The descriptive statistics are reported in the "Pooled" columns of Table 1 and are as follows. Approximately one quarter of the sample is in the blended course and approximately one half of the sample had taken a principles course prior to the semester of data collection. The average number of credits earned by students in the sample is approximately 40, indicating that the average student in the sample is in his/her sophomore year. In addition, approximately two thirds of the sample is male. Further, approximately 94 percent of the sample enrolled in 2009, 2010, and 2011, with the majority of the students enrolling in 2010. Finally, nearly 60 percent of the students in the sample are business or pre-business students; this is unsurprising, since both introductory macroeconomics and microeconomics are core requirements for the business school. Approximately 33% of the sample is in the school of arts and sciences and small percentages of students are in the schools of engineering, nursing and visual and performing arts.

**[Table 1 Approximately Here]**

Table 1 also reports the means of the variables of interest separately for the blended sections and F2F sections. The "Diff." column reports the difference in means between the blended and F2F sections and the "T-Stat" column reports the results of the t-test testing whether the difference in means is significantly different from zero assuming unequal variances. The results suggest that students in F2F sections performed significantly better on the exam overall, on the short answer section, on the SLO specific short answer questions and on the SLO specific online homework assignment. On average, fewer students in the blended group have taken principles of economics before and these students have significantly fewer cumulative credits. These two significant differences are likely because all blended sections took place in the Fall 2012 semester and a majority of the observations from the F2F sections came from the following semester. The proportion of students who are U.S. citizens is 10% lower in the blended sections than in the F2F sections and the proportion of students who enrolled in 2011 is lower in the blended section, suggesting that first-year students are less likely to enroll in a blended course.

## ESTIMATION STRATEGY

While the significant differences in the means of the outcomes may suggest that blending has a negative effect on learning outcomes, these effects may be moderated by other factors. In the following sections we present two techniques to account for individual characteristics. First, we estimate a series of OLS regressions, using a dummy variable for blended status as the key variable of interest. While OLS can control for individual characteristics, if individuals select non-randomly into blended or F2F courses, then the estimate of the effect of blending will be biased. Our second technique estimates the propensity score, the probability that a student chooses to enroll in a blended course, and matches individuals based on their propensity score to capture the true treatment effect.

### Benchmark OLS Regressions

To account for the aforementioned individual characteristics, we estimate a simple OLS regression to determine the effects of blending on student learning; if students select into classes randomly, then the OLS estimates, if specified correctly, will provide an unbiased estimate of the treatment effect. The OLS specification is as follows:

$$y_i = \beta_0 + \gamma BLENDED_i + \boldsymbol{CONTROLS_i}\beta_{CONTROLS} \quad (1)$$

where $y_i$ denotes student $i$'s performance on one of the assessments of interest, $BLENDED_i$ is an indicator variable taking a value of 1 if individual $i$ was enrolled in a blended course, and $\boldsymbol{CONTROLS}_i$ is a matrix of variables that control for student performance on the assessments. However, if selection into the blended sections is non-random, then the estimate of the treatment effect is biased.

### Propensity Score Matching

Selection into a blended course may be non-random for a number of reasons. It is important to note that the blended sections included in this study were not advertised as blended courses. That is, a student enrolling in the course did not know the course was blended until the first day of class. As a result, we do not expect to find as pronounced a selection bias as previously found with fully online courses (Coates et al. 2004). That said, students have the option to add and drop classes in the first week of the semester, so students who learn that a course is blended may drop the course because they do not wish to complete online work, or students may choose to enroll in a blended course because of the flexibility of completing some of the coursework online. Additionally, students who are more comfortable with the course or with their academics in general may be more likely to enroll in a blended course, given the element of independent work. Other reasons that students may non-randomly select into a particular section of a course include the time and days of the week that the course meets and the professor. If these reasons apply, the effect estimated in Equation 1 will not capture the true treatment effect.

To account for the aforementioned sources of bias, we employ a propensity score matching (PSM) approach. This section contains a brief description of the model and its assumptions. A more detailed explanation can be found in Rosenbaum and Rubin (1983), Dehejia and Wahba (1999), Dehejia and Wahba (2002), Wooldridge (2002) and Caliendo and Kopeinig (2008). The values of interest are the average treatment effect (ATE) and the average treatment effect on the treated (ATT). Following conventional notation, let $y_{1i}$ be the outcome of interest for individual $i$ if the individual is subjected to treatment (enrollment in a blended class). Let $y_{0i}$ be the outcome of individual $i$ without treatment (enrollment in a F2F class). Further, let $w_i$ be a treatment indicator, taking a value of one if individual $i$ has received treatment and zero otherwise. The ATE is the average difference in outcomes,

(2) $$ATE \equiv E(y_1 - y_0),$$

and the ATT can be defined as the average difference in outcomes, conditional on treatment:

(3) $$ATT \equiv E(y_1 - y_0 | w = 1)$$

Note that the econometrician only observes $y_{1i}$ or $y_{0i}$ for each individual $i$. As a result, to obtain the true value of the ATT, it is necessary include additional assumptions. In a detailed discussion of this model, Wooldridge (2002) notes that if $y_1$ and $y_0$ are independent of treatment status, then the expression for the ATT reduces to the difference in mean outcomes between individuals who were enrolled in a blended course and those who were not.

To account for the selection into treatment, we estimate the propensity score and determine the ATT by matching individuals based on its value. The propensity score is simply the probability of treatment based given a number of covariates, $\boldsymbol{X}$.[4] The propensity score given the covariates in $\boldsymbol{X}$ is given by:

(4) $$p(\boldsymbol{X}) \equiv \Pr(w = 1 | \boldsymbol{X})$$

[4] Caliendo and Kopeinig (2008) discusses ways to estimate the propensity score and reports that for binary treatments, there is little difference between a probit specification and a logit specification. We use a probit specification in what follows.

The following section describes the estimation of the propensity score and the methods used to compute the ATT. Additional details about the assumptions required for estimating the ATT and evidence that the assumptions hold are available from the authors upon request.[5]

**[Table 2 Approximately Here]**

### Propensity Score Estimation and Matching Algorithm Choice

The propensity score specification is provided in Table 2. The specification includes variables that may affect a student's selection of a blended or F2F course. The specification includes whether the student has taken principles of economics before, the number of cumulative credits earned, cumulative GPA, SAT math and verbal scores, a citizenship indicator, a gender indicator, a white/non-white indicator, an instructor indicator, an interaction term between teacher and citizenship status[6], indicators reporting whether the student enrolled for the first time in 2008, 2009, 2010 or 2011, an indicator variable reporting if the student is enrolled in the business school, and an indicator variable reporting whether the student is enrolled in the school of arts and sciences.

One of the assumptions of PSM is that after individuals are matched based on the propensity score, the treatment group and the control group should be nearly identical with respect to the variables used to estimate the propensity score. One way to verify this assumption is to perform a t-test of means across treatment status for each control variable in the matched sample. Table 3 reports the results of these t-tests. For each of the control variables, there is no significant difference in means between the blended and F2F sections; for the outcomes, only online homework retained a significant, negative effect of blending.[7]

**[Table 3 Approximately Here]**

In order to compute the ATE and the ATT, the econometrician must decide what algorithm to use in order to match treated and untreated individuals based on their propensity scores. Caliendo and Kopeinig (2008) present

[5] Propensity score matching requires two key assumptions in order to compute the correct treatment effcts and are described in J. J. Heckman, Ichimura, and Todd (1997). The assumptions are the strong ignorability of treatment and balancing property. Before estimating the treatment effects, which are reported in the next section, we checked whether these assumptions hold and found strong evidence that both assumptions are satisfied. More details about the assumption and the tests performed are available from the authors upon request.
[6] Because one of the instructors had a disproportionately large number of non-US citizens enrolled in his section, we include an interaction term between instructor and citizenship status. This term is included to satisfy the balancing property to make the treatment and control groups in the matched sample look similar. This interaction is included because Instructor 2 had a relatively large proportion of students who were not US citizens in the blended section; accounting for this balances the treatment and control groups.
[7] To perform this test, we used the user-written STATA program, "PSMATCH2" (Leuven and Sianesi 2003)

several common matching techniques and discuss the benefits and drawbacks of each technique. For each technique, there is a tradeoff between the bias and the efficiency of the estimates. In the results section, we compare the treatment effects across a number of different matching specifications.

When we determine the treatment effects, we consider a number of different estimation techniques. The first technique used is the nearest neighbor (NN) matching algorithm. Nearest neighbor matching with replacement, a maximum allowable caliper and one-to-one matching is the method with the lowest bias and lowest efficiency (Caliendo and Kopeinig 2008). This algorithm matches individuals in the treatment group to an individual in the control group whose propensity score is closest to that of the treated individual within the specified caliper.[8] We use the NN algorithm with replacement, allowing one individual in the control group to match up with several individuals in the treatment group, thus further reducing the bias of the estimates.[9]

Second, we use a nearest neighbor algorithm based on a Mahalanobis metric.[10] One drawback of the NN algorithm as well as other conventional matching algorithms is that it has small bias, but large standard errors, especially in small samples (Zhao 2004). Zhao (2004) suggests that matching based on the Mahalanobis metric tends to be more robust to small sample size.[11] Third, we estimate the treatment effect using a kernel matching technique based on a Gaussian kernel, a widely used matching technique.[12] Finally, we estimate the treatment effects for each outcome using radius matching with a maximum allowable caliper, another popular matching technique (Caliendo and Kopeinig 2008). For each of the matching algorithms, the standard errors are obtained by a bootstrapping procedure with 200 replications, following Lechner (2002).

---

[8] The econometrician must also decide the caliper size. Smith and Todd (2005) reports that there is no way to know *a priori* the appropriate caliper size. Following Rosenbaum and Rubin (1985) we choose a caliper size to be one quarter of the standard deviation of the propensity score; the caliper sizes range is approximately 5 percent. The effects are robust to a number of different caliper values; results of this robustness check are available upon request.

[9] Researchers have modified the NN algorithm to allow each treated observation to be matched up with multiple control observations. However, since this increases the bias of the estimated effects, we impose one-to-one matching (Caliendo and Kopeinig 2008).

[10] The Mahalanobis metric is an alternate way to measure distance between treated and non-treated observations. A more detailed analysis of this method can be found in Rubin (1980).

[11] This metric has also been examined in recent papers in the context of matching techniques (Abadie and Imbens 2006; (Imbens 2004).

[12] There is an extensive literature that examines the benefits and drawbacks of various matching algorithms. We refer the reader to LaLonde (1986), Heckman, Ichimura, and Todd (1997), Heckman, Ichimura, and Todd (1998), Lechner (2002), Wooldridge (2002), Zhao (2004), Smith and Todd (2005) and Caliendo and Kopeinig (2008).

## RESULTS

### OLS Estimates

Table 4 reports the results from the OLS specification described in the previous section (Equation (1)). Columns (1) through (3) of Table 4 reports the estimated coefficients for the models using the overall (not SLO specific) assessments as the dependent variables: the exam score, the score on the multiple choice section of the exam and the score on the short answer section of the exam, respectively. Columns (4) through (6) of Table 4 reports the coefficient estimates using the SLO specific assessments as the dependent variables: the scores on the SLO specific short answer questions, the percent of the SLO specific multiple choice questions answered correctly, and the scores on the online homework pertaining to comparative advantage and opportunity cost, respectively. Robust standard errors are provided below each estimate in parentheses.

**[Table 4 Approximately Here]**

Including the individual controls in Table 4 causes the effect of blending to become insignificant. For all outcomes, the addition of the controls results in a marked decrease in the difference between blended and F2F sections. The only outcome for which the effect of blending is almost significant at the 5% level is the online homework; the results indicate that blended coursework is associated with a lower online homework score of approximately four points. In addition to the attenuated effects of blending, several individual characteristics are associated with improved outcomes. First, there is a significant, positive effect for individuals who have taken principles of economics before; for exam scores, prior enrollment in a principles course is associated with nearly a six point increase in the overall exam score. However, prior enrollment in a principles course does not have a significant effect on the SLO specific outcomes. A likely explanation for this is that the other topic covered on the exam was the supply/demand model, and students who have taken principles before have been exposed to this topic throughout their first principles course. By contrast, comparative advantage and opportunity cost is only covered at the beginning of the course. Second, a higher cumulative GPA is associated with higher scores both overall and on the multiple choice section. Finally, unsurprisingly, there is a large, significant effect of both SAT scores on student outcomes; a 100 point increase in SAT scores (either math or verbal) is associated with a 5 point increase in the total exam score.[13]

**Propensity Score Results**

Table 5 reports the ATE computed by the propensity score methods described in the previous section for each outcome of interest. Column (1) reports the coefficients on blended status for the OLS regressions. Columns (2), (3) (4) and (5) report the ATE estimates for the nearest neighbor matching, the nearest neighbor matching with the Mahalanobis metric, the kernel matching estimator, and the radius matching estimator, respectively. Bootstrapped standard errors are provided below each estimate in parentheses.

**[Table 5 Approximately Here]**

With a few exceptions, the results indicate no significant effects of blending. The exceptions are as follows. The nearest neighbor matching algorithm (column 2) produces significant, negative effects of blending both on the short answer section and on the online homework assignment; on average, blended students earned 5 points less on the short answer questions and 6 points left on the online homework assignment. There is a significant, positive effect of blending on the SLO-specific multiple choice questions; according to column (2), students in the blended section answered one more SLO-specific multiple choice question correctly.[14] One possible explanation for this positive effect is that the low-stakes online reading quiz consisted entirely of multiple choice questions. Unlike students in the F2F sections, students in the blended classes were exposed to multiple choice questions similar to those seen on the exam. As a result, it is unsurprising that the blended section performed relatively well on those sections.

In addition to the significant values for the ATE, the results suggest the existence of selection bias in the OLS estimates, especially when compared with the NN algorithm. With the exception of the SLO specific multiple choice questions, the OLS estimates understate the effect of blending. For example, the OLS estimate of the effect of blending on the total exam score is -0.65 points, whereas the PSM estimate is -5.5 points. These results highlight the importance accounting for selection bias when determining the effects of blended coursework on student learning.

[13] In addition to the significant effects, it should also be noted that there does not seem to be any effect of instructor on student outcomes. Further, adding the instructor indicator variable has negligible effects on the effects of blending for all outcomes.
[14] This significant effect also exists using the kernel and radius matching algorithms, though they are not as strong.

A more widely-used measure of the treatment effect is the ATT. The ATT measures the effect of treatment on those in the treatment group, and may be a more relevant measure of the treatment effect (Heckman 1997). In our case, the ATT measures the average difference in outcomes between students in blended classes and the counterfactual outcomes they would have achieved in a F2F class. Table 6 reports these results. The results suggest that there is no significant effect of blending on student learning outcomes, regardless either of the outcome or of the method used to compute the treatment effect. However, this may be due to sample size; since the sample size is small, the standard errors are relatively large. As a result, in addition to statistical significance, we also focus on the magnitudes of the insignificant effects to determine if the treatment effect is substantively significant.

The magnitudes of the ATTs are also small. The effects vary based on estimation technique; in what follows, we present the range of values of the effect of blending. For exam scores, the effect of blending ranges from a decrease of 0.6 points to an increase of 2.8 points (the exam was out of 100 points). For the multiple choice section, the effect of blending is small and positive, ranging from 0.16 points to 2.4 points.[15] For the short answer section, the effect of blending ranges from a decrease of 1.6 points to an increase of 0.8 points.[16] For the SLO specific short answer questions, the effect of blending ranges from a decrease of 0.75 points to an increase of 0.9 points. The SLO specific short answer question had ten parts, each ranging from 2 to 4 points in value, so a decrease of one point is half of one part of the question. For the percentage of SLO-specific multiple choice questions, the effect ranges from 6 percent to 14 percent; however, there were five SLO specific multiple choice questions on the exam, so an increase of 14 percent amounts to less than one question. For the online homework, the effect of blending results in a decrease in scores ranging from 1.3 to 4 points (out of 50 points).

## DISCUSSION AND CONCLUSION

Few studies examine the effects of blended learning on student outcomes in economics, despite the growing number of blended course offerings among universities. Thus, this study fills a gap in the literature by being one of the few studies to examine blending in the context of economics courses, and being the only one, to our knowledge, to account for the selection into a blended course.

[15] The multiple choice section was out of 50 points, and each question was worth 2.5 points. The estimated effect of blending is less than one correct multiple choice question.
[16] The short answer section was out of 50 points.

Our results, with few exceptions, suggest no significant effect of blending on any of the outcomes considered. Both for the general learning objectives and for the specific learning objectives, significant differences in the raw means exist, but after controlling for individual academic and demographic characteristics, we find no significant difference in outcomes between the blended and non-blended sections. Both the OLS specification and the PSM specifications produce this result, suggesting no causal link between blended coursework and reduced learning.

The only outcome for which the effect of blending is consistently negative is the SLO-specific online homework assignment. Depending on the specification, the ATT of online homework scores ranges from a decrease of 4 points to a decrease of 1.3 points. This effect is considerably larger than the others; one possible reason for this result is that students in blended classes had other assignments besides the online homework to which they allocated their time. While students in the blended sections were less successful on the online homework, they learned the material using the wiki and discussion boards, so there was no overall loss of learning. In addition, students in the F2F sections had one more class period *before* the online assignment was due. As a result, the negative effect of blending on this outcome may be an artifact of having less exposure to the material; this negative effect disappeared when students took the exam.

Although the results presented here are not generalizable to the population of college students, our results provide a number of implications for blended coursework and questions for future research. First, the results suggest that blended coursework in principles of economics may provide a flexible alternative toF2F classes without sacrificing student learning. Second, the results suggest that like fully online courses, students are selecting into, and out of blended classes, and accounting for this selection is important when determining the effects of blended coursework. Subsequent research on this topic should be careful to account for this type of selection when drawing conclusions about the effects of blended classes on student learning. Further, while the results suggest no loss of student learning, an important area for further investigation is determining the costs involved in delivering a blended course compared to a traditional course. In addition to the sunk cost of developing the blended course, there may be an additional time cost of monitoring and facilitating the online components of the course. Given the recent increase in blended course offerings, a careful analysis of these costs is warranted.

## REFERENCES


Abadie, Alberto, and Guido W. Imbens. 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica* 74 (1) (January): 235–267. doi:10.1111/j.1468-0262.2006.00655.x.

Allen, I. Elaine, Jeff Seaman, and Richard Garrett. 2007. *Blending In: The Extent and Promise of Blended Education in the United States*. The Sloan Consortium. http://sloanconsortium.org/publications/survey/pdf/Blending_In.pdf.

Anstine, Jeff, and Mark Skidmore. 2005. "A Small Sample Study of Traditional and Online Courses with Sample Selection Adjustment." *The Journal of Economic Education* 36 (2) (April 1): 107–127.

Arbaugh, J.B., Michael R. Godfrey, Marianne Johnson, Birgit Leisen Pollack, Bruce Niendorf, and William Wresch. 2009. "Research in Online and Blended Learning in the Business Disciplines: Key Findings and Possible Future Directions." *The Internet and Higher Education* 12 (2) (June): 71–87. doi:10.1016/j.iheduc.2009.06.006.

Brown, Byron W, and Carl E Liedholm. 2002. "Can Web Courses Replace the Classroom in Principles of Microeconomics?" *American Economic Review* 92 (2) (May): 444–448. doi:10.1257/000282802320191778.

Caliendo, Marco, and Sabine Kopeinig. 2008. "SOME PRACTICAL GUIDANCE FOR THE IMPLEMENTATION OF PROPENSITY SCORE MATCHING." *Journal of Economic Surveys* 22 (1) (February): 31–72. doi:10.1111/j.1467-6419.2007.00527.x.

Coates, Dennis, Brad R. Humphreys, John Kane, and Michelle A. Vachris. 2004. "``No Significant Distance'' Between Face-to-face and Online Instruction: Evidence from Principles of Economics." *Economics of Education Review* 23 (3): 533–546.

Dehejia, Rajeev H., and Sadek Wahba. 1999. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94 (448): 1053–1062.

———. 2002. "Propensity Score-Matching Methods For Nonexperimental Causal Studies." *The Review of Economics and Statistics* 84 (1): 151–161.

Garnham, Carla, and Robert Kaleta. 2002. "Introduction to Hybrid Courses." *Teaching With Technology Today* 8 (6) (March 20). http://www.wisconsin.edu/ttt/articles/garnham.htm.

Garrison, Randy, Heather Kanuka, and David Hawes. 2002. "Blended Learning: Archetypes for More Effective Undergraduate Learning Experiences". University of Calgary: Learning Commons.

Heckman, James. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *The Journal of Human Resources* 32 (3) (July 1): 441–462.

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *The Review of Economic Studies* 64 (4) (October): 605. doi:10.2307/2971733.

———. 1998. "Matching As An Econometric Evaluation Estimator." *Review of Economic Studies* 65 (2) (April): 261–294. doi:10.1111/1467-937X.00044.

Imbens, Guido W. 2003. "Sensitivity to Exogeneity Assumptions in Program Evaluation." *American Economic Review* 93 (2) (May): 126–132. doi:10.1257/000282803321946921.

———. 2004. "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review." *Review of Economics and Statistics* 86 (1) (February): 4–29. doi:10.1162/003465304323023651.

Kelley, T. L. 1939. "The Selection of Upper and Lower Groups for the Validation of Test Items." *Journal of Educational Psychology* 30 (1): 17–24. doi:10.1037/h0057123.

LaLonde, Robert J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *The American Economic Review* 76 (4) (September): 604–620.

Lechner, Michael. 2002. "Some Practical Issues in the Evaluation of Heterogeneous Labour Market Programmes by Matching Methods." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 165 (1) (February): 59–82. doi:10.1111/1467-985X.0asp2.

Leuven, Edwin, and Barbara Sianesi. 2003. "PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing." *Statistical Software Components S432001, Boston College Department of Economics, Revised 13 Dec 2011.*

Means, Barbara, Yuki Toyama, Robert Murphy, Marianne Bakia, and Karla Jones. 2009. "Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies." *Structure* 15 (20): 94.

Navarro, Peter, and Judy Shoemaker. 2000. "Performance and Perceptions of Distance Learners in Cyberspace." *American Journal of Distance Education* 14 (2) (January): 15–35. doi:10.1080/08923640009527052.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41–55. doi:10.1093/biomet/70.1.41.

———. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score." *The American Statistician* 39 (1) (February): 33. doi:10.2307/2683903.

Rubin, Donald B. 1980. "Bias Reduction Using Mahalanobis-Metric Matching." *Biometrics* 36 (2) (June): 293–298.

Smith, Jeffrey A., and Petra E. Todd. 2005. "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics* 125 (1-2) (March): 305–353. doi:10.1016/j.jeconom.2004.04.011.

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

Vaughan, Norm. 2007. "Perspectives on Blended Learning in Higher Education." *International Journal on E-Learning* 6 (1) (January): 81–94.

Williams, Christina. 2002. "Learning On-line: A Review of Recent Literature in a Rapidly Expanding Field." *Journal of Further and Higher Education* 26 (3) (August): 263–272. doi:10.1080/03098770220149620.

Wooldridge, Jeffrey M. 2002. *Econometric analysis of cross section and panel data*. Cambridge, Mass.: MIT Press.

Zhao, Zhong. 2004. "Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence." *Review of Economics and Statistics* 86 (1) (February): 91–107. doi:10.1162/003465304323023705.

## Appendix A. Tables

**Table 1: Descriptive Statistics and T-Tests of Means**

| | Pooled | | Blended | | Face-to-Face | | T-Test of Means | |
|---|---|---|---|---|---|---|---|---|
| Variable | Mean | S.D. | Mean | S.D. | Mean | S.D. | Diff.[a] | T-Stat |
| **Outcomes** | | | | | | | | |
| *Exam Score* | 78.20 | 13.55 | 75.52 | 13.37 | 79.13 | 13.52 | -3.61 | -2.10* |
| *Mult. Choice Score* | 37.70 | 6.52 | 37.44 | 6.37 | 37.80 | 6.59 | -0.36 | -0.43 |
| *Short Answer* | 40.49 | 8.86 | 38.08 | 9.05 | 41.33 | 8.66 | -3.25 | -2.83** |
| *MC Percent (SLO)* | 45.66 | 37.97 | 47.32 | 40.19 | 45.08 | 37.24 | 2.23 | 0.44 |
| *Short Answer (SLO)* | 18.56 | 5.26 | 17.01 | 5.35 | 19.09 | 5.14 | -2.08 | -3.06** |
| *Online Homework* | 31.59 | 16.11 | 27.85 | 16.59 | 32.88 | 15.78 | -5.04 | -2.40* |
| **Controls** | | | | | | | | |
| *Blended* | 0.26 | 0.44 | 1.00 | 0.00 | 0.00 | 0.00 | 1 | . |
| *Taken Principles* | 0.51 | 0.50 | 0.33 | 0.47 | 0.57 | 0.50 | -0.24 | -3.95*** |
| *Cum. Credits* | 39.88 | 20.22 | 34.39 | 17.77 | 41.79 | 20.70 | -7.4 | -3.11** |
| *Cum. GPA* | 2.59 | 0.83 | 2.46 | 0.92 | 2.64 | 0.79 | -0.18 | -1.55 |
| *SAT-Math* | 524.84 | 79.92 | 511.34 | 80.70 | 529.53 | 79.28 | -18.19 | -1.77 |
| *SAT-Verb* | 486.89 | 72.49 | 480.24 | 76.19 | 489.19 | 71.18 | -8.95 | -0.93 |
| *U.S. Citizen* | 0.93 | 0.26 | 0.85 | 0.36 | 0.95 | 0.21 | -0.1 | -2.40* |
| *Male (1 = Male)* | 0.66 | 0.47 | 0.66 | 0.48 | 0.67 | 0.47 | -0.01 | -0.11 |
| *Instructor 1* | 0.43 | 0.50 | 0.59 | 0.50 | 0.38 | 0.49 | 0.2 | 3.25** |
| **Race/Ethnicity** | | | | | | | | |
| *Asian* | 0.06 | 0.23 | 0.09 | 0.28 | 0.05 | 0.21 | 0.04 | 1.14 |
| *Black* | 0.09 | 0.29 | 0.10 | 0.30 | 0.09 | 0.29 | 0 | 0.11 |
| *Cape Verdean* | 0.02 | 0.15 | 0.02 | 0.16 | 0.02 | 0.14 | 0 | 0.16 |
| *Hispanic* | 0.05 | 0.22 | 0.06 | 0.24 | 0.05 | 0.21 | 0.01 | 0.48 |
| *Non-Resident Alien* | 0.02 | 0.12 | 0.02 | 0.16 | 0.01 | 0.11 | 0.01 | 0.63 |
| *Not Specified* | 0.01 | 0.11 | 0.01 | 0.11 | 0.01 | 0.11 | 0 | -0.04 |
| *More than two races* | 0.04 | 0.21 | 0.05 | 0.22 | 0.04 | 0.20 | 0.01 | 0.23 |
| *White* | 0.70 | 0.46 | 0.65 | 0.48 | 0.72 | 0.45 | -0.08 | -1.29 |
| **1st Year Enrolled** | | | | | | | | |
| *2007* | 0.01 | 0.11 | 0.02 | 0.16 | 0.01 | 0.09 | 0.02 | 0.88 |
| *2008* | 0.04 | 0.21 | 0.04 | 0.19 | 0.05 | 0.21 | -0.01 | -0.4 |
| *2009* | 0.20 | 0.40 | 0.26 | 0.44 | 0.18 | 0.39 | 0.07 | 1.35 |
| *2010* | 0.55 | 0.50 | 0.60 | 0.49 | 0.54 | 0.50 | 0.06 | 0.94 |
| *2011* | 0.18 | 0.39 | 0.09 | 0.28 | 0.22 | 0.41 | -0.13 | -3.19** |
| *2012* | 0.01 | 0.08 | 0.00 | 0.00 | 0.01 | 0.09 | -0.01 | -1.42 |
| **Program of Major** | | | | | | | | |
| *Arts and Sciences* | 0.33 | 0.47 | 0.23 | 0.42 | 0.36 | 0.48 | -0.13 | -2.35* |
| *Business* | 0.59 | 0.49 | 0.68 | 0.22 | 0.56 | 0.24 | 0.12 | 1.96 |
| *Engineering* | 0.06 | 0.24 | 0.05 | 0.47 | 0.06 | 0.50 | 0.01 | 0.51 |
| *Nursing* | 0.01 | 0.11 | 0.02 | 0.16 | 0.01 | 0.09 | 0.02 | 0.88 |
| *Visual/Perf. Arts* | 0.00 | 0.06 | 0.01 | 0.11 | 0.00 | 0.00 | 0.01 | 1 |
| *N* | 318 | | 82 | | 236 | | | |

[a] This column reports the difference in means between the blended and non-blended sections

* p< .05; ** p < .01; *** p < .001

**Table 2: Propensity Score Specification-Probit Results**

| Variable | Coef. (S.E). |
|---|---|
| Taken Econ. Principles | -0.676** |
| | (0.235) |
| Cumulative Credits | -0.027*** |
| | (0.008) |
| Cumulative GPA | -0.193 |
| | (0.123) |
| SAT Math | 0 |
| | (0.001) |
| SAT Verbal | 0.001 |
| | (0.001) |
| US Citizen | -1.713*** |
| | (0.453) |
| Male | -0.037 |
| | (0.197) |
| White | -0.106 |
| | (0.217) |
| Instructor 1 | -1.019 |
| | (0.627) |
| Enrolled in 2008 | 1.307 |
| | (0.778) |
| Enrolled in 2009 | 0.606 |
| | (0.605) |
| Enrolled in 2010 | -0.328 |
| | (0.592) |
| Enrolled in 2011 | -1.728** |
| | (0.636) |
| Instructor 1 x Citizen | 1.744** |
| | (0.645) |
| Business | 0.729* |
| | (0.366) |
| Arts & Sciences | -0.172 |
| | (0.368) |
| Constant | 2.006* |
| | (1.015) |
| Pseudo R-Square | 0.258 |
| P-Value ($H_0$: All Coefs. Sig.) | < 0.001 |
| *N* | 318 |

Results from the probit estimation of the propensity score are presented here. Standard errors are provided below each estimate in parentheses.
* p< .05; ** p < .01; *** p < .001

**Table 3: Comparison of the Matched and Unmatched Samples**

| | Unmatched Sample | | | Matched Sample | | |
|---|---|---|---|---|---|---|
| | Blended | F2F | T-Stat | Blended | F2F | T-Stat |
| Exam Score | 75.52 | 78.20 | -2.10* | 75.70 | 75.87 | -0.07 |
| Mult. Choice Score | 37.44 | 37.70 | -0.43 | 37.47 | 37.31 | 0.14 |
| Short Answer Score | 38.08 | 40.49 | -2.83** | 38.23 | 38.56 | -0.23 |
| % MC Correct (SLO) | 47.32 | 45.66 | 0.44 | 16.94 | 16.05 | 0.95 |
| Short Answer (SLO) | 17.01 | 18.56 | -3.06** | 48.61 | 38.48 | 1.70 |
| Online HW | 27.85 | 31.59 | -2.40* | 27.43 | 29.92 | -0.93 |
| Taken Principles | 0.51 | 0.33 | -3.95*** | 0.34 | 0.22 | 1.78 |
| Cum. Credits | 39.88 | 34.39 | -3.11** | 34.68 | 31.71 | 1.13 |
| Cum. GPA | 2.59 | 2.46 | -1.55 | 2.45 | 2.35 | 0.65 |
| SAT-Math | 524.84 | 511.34 | -1.77 | 509.37 | 529.49 | -1.58 |
| SAT-Verb | 486.89 | 480.24 | -0.93 | 481.39 | 487.47 | -0.54 |
| U.S. Citizen | 0.93 | 0.85 | -2.40* | 0.86 | 0.95 | -1.91 |
| Male (1 = Male) | 0.66 | 0.66 | -0.11 | 0.66 | 0.78 | -1.78 |
| Instructor 1 | 0.43 | 0.59 | 3.25** | 0.58 | 0.70 | -1.49 |
| White | 0.65 | 0.72 | -1.34 | 0.66 | 0.78 | -1.78 |
| Enrolled in 2008 | 0.04 | 0.05 | -0.38 | 0.04 | 0.00 | 1.75 |
| Enrolled in 2009 | 0.26 | 0.18 | 1.44 | 0.27 | 0.14 | 1.99* |
| Enrolled in 2010 | 0.60 | 0.54 | 0.93 | 0.58 | 0.72 | -1.85 |
| Enrolled in 2011 | 0.09 | 0.22 | -2.66** | 0.09 | 0.08 | 0.29 |
| Arts & Sciences | 0.68 | 0.56 | 1.9 | 0.68 | 0.73 | -0.70 |
| Business | 0.23 | 0.36 | -2.21* | 0.24 | 0.22 | 0.38 |

The columns labeled "T-Stat." report the t-statistic of the null hypothesis of no difference in means between the blended and non-blended sections, assuming unequal variances.

* p < .05; ** p < .01; *** p < .001

**Table 4: OLS Results**[a]

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | General Learning Outcomes | | | Specific Learning Outcomes | | |
| | Exam | M.C. | S.A. | S.A. | M.C. % | Online HW |
| Blended Section | -0.646 | 1.024 | -1.661 | -0.746 | 6.567 | -4.004 |
| | (1.646) | (0.832) | (1.168) | (0.678) | (5.933) | (2.165) |
| Taken Econ. Principles | 5.601** | 3.764*** | 1.801 | 0.265 | 5.665 | -1.636 |
| | (1.702) | (0.870) | (1.161) | (0.682) | (5.882) | (2.418) |
| Cumulative Credits | 0.018 | -0.01 | 0.028 | 0.032 | -0.059 | 0.215** |
| | (0.048) | (0.024) | (0.037) | (0.023) | (0.185) | (0.073) |
| Cumulative GPA | 1.774 | 1.258** | 0.513 | -0.024 | 2.659 | 1.802 |
| | (0.920) | (0.433) | (0.627) | (0.390) | (3.009) | (1.211) |
| SAT Math | 0.051*** | 0.020*** | 0.031*** | 0.023*** | 0.014 | 0.038** |
| | (0.010) | (0.005) | (0.007) | (0.004) | (0.031) | (0.014) |
| SAT Verbal | 0.050*** | 0.025*** | 0.024*** | 0.009* | 0.03 | -0.002 |
| | (0.010) | (0.005) | (0.007) | (0.004) | (0.036) | (0.016) |
| US Citizen | -2.316 | -1.998 | -0.315 | 1.224 | -6.238 | -7.759 |
| | (4.024) | (1.914) | (2.765) | (1.646) | (12.605) | (5.128) |
| Male | 1.095 | 0.665 | 0.417 | -0.193 | 1.372 | 2.432 |
| | (1.437) | (0.676) | (1.010) | (0.562) | (4.954) | (1.846) |
| White | 1.835 | 0.934 | 0.889 | 0.658 | 5.599 | 1.158 |
| | (1.608) | (0.765) | (1.092) | (0.608) | (5.212) | (2.213) |
| Instructor 1 | 6.087 | 3.262 | 2.8 | 2.229 | -15.263 | 0.768 |
| | (5.594) | (2.347) | (4.281) | (2.025) | (17.204) | (5.076) |
| Instructor 1 x Citizenship | -4.225 | -1.399 | -2.831 | -3.471 | 13.152 | 1.436 |
| | (5.558) | (2.372) | (4.249) | (2.104) | (17.380) | (5.445) |
| Enrolled in 2008 | 6.565 | 3.206 | 3.346 | 3.673 | 12.247 | -6.4 |
| | (5.440) | (3.120) | (4.714) | (3.212) | (22.299) | (8.089) |
| Enrolled in 2009 | 5.002 | 1.696 | 3.314 | 4.465 | -0.445 | 3.53 |
| | (4.807) | (2.870) | (3.843) | (2.652) | (18.634) | (7.623) |
| Enrolled in 2010 | 4.27 | 1.479 | 2.795 | 4.918 | -5.206 | 5.504 |
| | (5.000) | (3.013) | (3.678) | (2.661) | (18.348) | (7.882) |
| Enrolled in 2011 | 8.624 | 4.04 | 4.604 | 6.194* | 8.075 | 10.381 |
| | (5.590) | (3.290) | (3.793) | (2.795) | (18.719) | (8.720) |
| Business | -1.37 | -1.555 | 0.196 | -0.801 | -13.811 | 3.709 |
| | (3.014) | (1.084) | (2.269) | (1.251) | (10.044) | (3.520) |
| Arts & Sciences | -3.24 | -2.188* | -1.062 | -1.317 | -8.834 | -2.364 |
| | (2.960) | (1.108) | (2.234) | (1.228) | (9.990) | (3.558) |
| Constant | 15.419 | 9.313* | 6.081 | -4.307 | 28.767 | -1.566 |
| | (9.082) | (4.467) | (6.303) | (3.879) | (29.371) | (11.326) |
| Adjusted R-Square | 0.297 | 0.293 | 0.192 | 0.246 | -0.008 | 0.116 |
| P-Value ($H_0$: All Coefs. Sig.) | < 0.001 | < 0.001 | < 0.001 | < 0.001 | 0.677 | < 0.001 |
| *N* | 318 | 318 | 318 | 318 | 318 | 318 |

[a] Robust standard errors are provided below each estimate in parentheses. The dependent variables for column (1) is the overall exam score; for column (2) the dependent variable is the score on the multiple choice section of the exam; for column (3) the dependent variable is score on the short answer section; for column (4) the dependent variable is the result on the short answer question pertaining to the SLO of interest; for column (5), dependent variable is the percentage of SLO specific multiple choice questions correctly answered; for column (6) the dependent variable is the score on the online homework assignment pertaining to the SLO.

* p< .05; ** p < .01; *** p < .001

**Table 5: Average Treatment Effects (*ATE*)**

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Outcome | OLS | NN (1) | NN (2) | Kernel | Radius |
| Exam Score | -0.646 | -5.495 | -2.22 | -1.258 | -0.97 |
| | (1.646) | (3.005) | (1.833) | (2.302) | (2.727) |
| Mult. Choice | 1.024 | -0.266 | 0.338 | 1.221 | 1.162 |
| | (0.832) | (1.200) | (0.911) | (1.138) | (1.176) |
| Short Answers | -1.661 | -5.223* | -2.552 | -2.472 | -2.125 |
| | (1.168) | (2.227) | (1.326) | (1.664) | (1.863) |
| Short Ans. (SLO) | -0.746 | -2.019 | -1.392 | -0.798 | -0.511 |
| | (0.678) | (1.069) | (0.735) | (0.873) | (0.969) |
| % MC Correct (SLO) | 6.567 | 19.613* | 7.925 | 13.976* | 15.724* |
| | (5.933) | (9.359) | (6.187) | (7.081) | (7.388) |
| Online HW | -4.004 | -6.241* | -3.586 | -2.45 | -2.406 |
| | (2.165) | (2.737) | (2.267) | (2.739) | (2.574) |
| N/# Matched Pairs | 318 | 60 | 60 | 318 | 318 |

Standard errors are provided below each estimate in parentheses. For the OLS models, robust standard errors are reported. For the PSM methods, the standard errors are computed using a bootstrap procedure with 200 replications. Column (1) reports the effects of blending estimated from Equation 1. Columns (2) - (5) report the ATE computed using propensity score matching procedures. Column (2) reports the results using nearest neighbor matching, with one-to-one matching, and with a maximum allowable caliper of 5.6%. Column (3) reports the results using nearest neighbor matching with one-to-one matching and a Mahalanobis metric. Column (4) reports the ATE computed using a kernel matching estimator with a Gaussian kernel. Finally, Column (5) reports the ATE computed using radius matching with a caliper size of 5.6%

* p < .05; ** p < .01; *** p < .001

**Table 6: Average Treatment Effects on the Treated (*ATT*)**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Outcome | NN (1) | NN (2) | Kernel | Radius |
| Exam Score | -0.177 | 2.805 | 1.192 | 1.881 |
| | (3.402) | (3.045) | (3.532) | (2.868) |
| Mult. Choice | 0.158 | 2.348 | 1.002 | 1.037 |
| | (1.516) | (1.29) | (1.582) | (1.296) |
| Short Answers | -0.335 | 0.457 | 0.191 | 0.844 |
| | (2.175) | (2.08) | (2.138) | (1.874) |
| Short Ans. (SLO) | 0.892 | 0.671 | 0.616 | 0.851 |
| | (1.325) | (1.297) | (1.246) | (1.222) |
| % MC Correct (SLO) | 10.127 | 11.951 | 10.385 | 13.678* |
| | (7.119) | (7.486) | (6.081) | (6.094) |
| Online HW | -2.487 | -1.299 | -2.776 | -2.64 |
| | (3.702) | (2.983) | (3.511) | (3.591) |
| *N*/# Matched Pairs | 60 | 60 | 318 | 318 |

Standard errors are provided below each estimate in parentheses. For the PSM methods, the standard errors are computed using a bootstrap procedure with 200 replications. This table reports the ATT computed using propensity score matching procedures. Column (1) reports the results using nearest neighbor matching, with one-to-one matching, and with a maximum allowable caliper of 5.6%. Column (2) reports the results using nearest neighbor matching with one-to-one matching and a Mahalanobis metric. Column (3) reports the ATT computed using a kernel matching estimator with a Gaussian kernel. Finally, Column (4) reports the ATT computed using radius matching with a caliper size of 5.6%

* p < .05; ** p < .01; *** p < .001