

Assessing the impact of blended learning on student performance¹

Do Won Kwak, Flavio M. Menezes and Carl Sherwood
The University of Queensland

October 30, 2013

¹We are thankful for the advice and technical support provided by the Evaluation Unit at the University of Queensland's Teaching and Educational Development Institute (TEDI). We are also thankful to Elle Parslow for research assistance.

Abstract

This paper assesses quantitatively the impact on student performance of a blended learning experiment within a large undergraduate first year course in statistics for business and economics students. We employ a differences-in-difference econometric approach, which controls for differences in student characteristics and course delivery method, to evaluate the impact of blended learning on student performance. Although students in the course manifest a preference for live lectures over online delivery, our empirical analysis strongly suggests that student performance is not affected (either positively or negatively) by the introduction of blended learning.

Key-words: Blended learning, assessment, quantitative analysis.

1 Introduction

This paper assesses an experiment with blended learning conducted at the University of Queensland. The experiment involved replacing a couple of two-hour traditional face to face lectures with a blended learning approach involving a one hour face to face lecture followed by material delivered exclusively online. Our key contribution is to measure the impact of the experiment on students' learning. We do so by comparing students' performances using two online quizzes, that cover the material delivered exclusively online, and four other quizzes that assess the material delivered in the traditional face to face lectures. In our comparison we control for various student characteristics such as gender, country of birth, program of study and final exam outcomes.

Our experiment is not in itself unique. The pervasiveness of online learning strategies as part of students' learning experiences in higher education is self-evident.¹ The two most prominent methods are online delivery (OL) and blended learning (BL). Online learning is where all the content is delivered online, with face to face contact limited to perhaps only tutorials. Blended learning includes a combination of materials that are delivered both face to face and online. Our contribution to the literature, however, is to provide additional, robust empirical evidence that, while students seem to be reluctant to entirely forgo the face to face experience, in reality their performance is not affected by the delivery method.

To be able to appreciate our results it is necessary to understand that there are many drivers that explain the take up of different teaching delivery methods that depart from the traditional, live, face to face (FtoF) lectures. These drivers include a combination of supply driven factors such as advances in technology that have reduced the cost to using online delivery methods, economies

¹In the US, for example, the number of students taking online classes increased from 1.6 million in 2002 to over 2.6 million in Fall 2005. (Allen and Seaman, 2006).

of scale and scope in the delivery of education to a large number of students (Twigg, 2013; Morris, 2008) and decreased funding per student (Mortensen, 2005). Drivers also include demand side factors such as different attitudes towards online delivery from the newer generation of students (Sebastianelli and Tamimi, 2011) and reduced student engagement with traditional face to face delivery (Exetera et al., 2010). While the introduction of online teaching may lower the costs to students to pursuing higher education, it may also provide weaker incentives for students to keep up with their studies as documented by Donovan et al. (2006).

These different drivers strongly suggest that pedagogy is only one of the main rationales for the introduction of online or blended learning. Therefore, consideration needs to be given on how to assess the success of online and blended learning. It should not be surprising then that a large body of literature has emerged on how to assess these novel teaching delivery strategies. By and large, there are two types of methods. The first strategy involves assessing the students' satisfaction with the delivery method. This is done usually through surveys or focus groups. The second strategy involves comparing student grades while employing various degrees of controls. Simple controls look to make use of the same lecturer or course materials, while more sophisticated approaches control for students' characteristics. In this paper, we pursue the first strategy to simply gauge students' attitudes and responses, and the second strategy to rigorously assess the impact of delivery method on student performance applying a range of controls. We present some selected references in Table 1 which follow both of these strategies.

It is obvious from Table 1 that there are no clear cut answers as to which strategy produces better learning outcomes. However, an emerging theme is that even when there are differences, either in students' preferences or perfor-

mance, such differences are often not substantial. For a more comprehensive list of references see, for example, <http://www.nosignificantdifference.org/> or US Department of Education (2009).

[Insert Table 1 here]

This paper is closely related to both Brown and Liedholm (2002) and Figlio et al. (2013). Brown and Liedholm (2002) assessed the impact of teaching mode on student performance in principles of microeconomics courses taught at Michigan State University in 2000 and 2001. The key question they examined was whether students enrolled in online courses learn more or less than students taught face to face or through blended learning. In addition, they sought to identify student characteristics, such as gender, race, university entry scores, or grade averages that were associated with better learning outcomes using one particular technology.

Our approach is similar in that we assess students' performance from blended learning vis-a-vis face to face lectures. The main difference is in the experiment design. While we design an experiment where all students are exposed to both face to face and blended learning – and we also compare to students' performances from previous cohorts who were only exposed to face to face teaching – in Brown and Liedholm (2002) students were assigned to one of three teaching modes. They then used percentage average marks from exam questions as the basis to draw conclusions for each mode of learning controlling for students' characteristics which included gender, race, course program, athletes, and honours student. As in our experiment, the textbook and course content did not change across delivery modes. Our experimental design, using a difference-in-difference (DID) method, is able to remove potential bias from self-selection of students between classes of blended learning and that of face to face. In particular, if self-selection into blended learning and student quality are correlated in

an unobserved manner, controlling by extensive students' characteristics variables is not enough to eliminate selection bias. However, the DID method can completely remove the selection bias.

Brown and Liedholm (2002) showed there was no significant difference in predicted scores across the three modes of instruction for definitional and recognition type questions. However, face to face students did significantly better than online students and better than blended learning students on the most complex material. In contrast, we assess students' performance using quizzes with similar level of difficulty or complexity in both face to face and blended learning modes of instruction. We present robust evidence that students' performance is not impacted when blended learning is used as a delivery mode and when we control for students' characteristics. Our results suggest that the way in which blended learning is implemented might be important for student performance. In particular, in our experiment the online content is used to support face to face content rather than be a substitute for it.

Figlio et al. (2013) reported on an experiment where controls were significantly more robust than those in the literature reviewed above. In particular, they reported on an experiment conducted in a large (1,600 to 2,600 students a semester) principles of microeconomics class taught by a single instructor at a large research intensive university. In this experiment students were randomly assigned to either an online or a live section for a course taught by one instructor and for which the support material such as web page, problem sets, teaching assistant support and exams were identical between the sections. The experiment was designed so that the only difference between the two treatments was the method of delivery of the lectures with some students viewing the lectures online and others face to face.

By comparing average grades across treatments, Figlio et al. (2013) showed

that students performed better in face to face than in online learning. This difference, however, was not significant. Adding controls, like gender, university entry scores, race and overall academic performance, leads to more precision (that is, smaller standard errors) but also a larger and statistically significant difference in average scores. These authors showed that when the controls are introduced, students' average scores are likely to be 2.5 higher (in a 100 point scale) under face to face teaching than under online teaching. These results suggest that online delivery, on its own, may have detrimental effects on students' learning. In contrast, we conceived our experiment to assess the impact of blended learning, where online delivery is part of a pedagogical approach that does not see face to face teaching replaced by online teaching. In our design, only two lectures are delivered through blended learning, which allows us to compare its impact on student performance within the same cohort of students. We also compare student performance with other cohorts who did not experience blended learning. As indicated above, we found robust evidence that blended learning does not adversely impact on student performance. Along with Figlio et al. (2013), this suggests that blended learning may lead to substantially better educational outcomes than online learning while delivering some of the benefits of online learning such as economies of scale and scope and cost savings.

The remainder of this paper is organised as follows. Section 2 describes the blended learning experiment and provides some qualitative information about the reaction and attitude of students to the introduction of blended learning. Section 3 presents the data, our empirical approach and results identifying the impact of the introduction of blended learning on students' performances. Finally, Section 4 discusses the results and provides some concluding remarks.

2 The blended learning experiment

Historically, the first year introductory statistics course for business and economics students at the University of Queensland has involved presenting thirteen weeks of lectures, with each lecture being two hours long and repeated twice a week. The lectures have always been delivered in a large face to face lecture environment, and over the last decade, been recorded for students to access after the lecture. Typically the lectures have used a combination of PowerPoint presentations, Excel demonstrations, and made use of a visualiser to work through hand calculations. Over the last three years, the assessment of students has involved six online quizzes, a mid-semester and final exam.

The experiment targeted lectures six and seven. These lectures were selected as they were in middle of the thirteen week course allowing students time to become familiar with various course learning activities. In particular, students would become familiar with the online quiz assessment requirements by the time they were exposed to quizzes three and four which were associated with the experiment in the course. This was considered important as the results from quizzes three and four were to be used as a key part of the experiment in the course.

The design of intervention lectures was based on various requirements. The main requirement was to reduce the face to face lecture time from two hours to one hour, with the second hour supplementing the one hour face to face lecture with online material accessed by students after the face to face lecture. The one hour lecture was to be designed to cover theoretical aspects, and to convey the “big picture” relevance of the lecture topic. The second hour of the traditional lecture was to be delivered online with the aim being to provide practical examples and application of theory covered in the face to face lecture. In other words, the online material was to build on the ideas presented in the face to face

lecture. In addition, the online material needed to be quickly and efficiently created from previously used lecture materials. The technology employed to create the online material from a lecturer's perspective needed to be low cost, easy to use, produce good quality visual and audio recordings, and be readily accessible to any lecturer. These requirements were considered essential to attract other staff to adopt this approach should research findings reveal benefits in student learning outcomes. From the student's perspective, it was decided all online materials must be accessible from various devices such as an iPad, iPhone, or laptop to provide maximum convenience and flexibility.

Based on the project design requirements, both the face to face and online materials were produced for lectures six and seven. Previously used course PowerPoint lecture slides were easily modified to present a one hour face to face lecture covering theoretical concepts. The online videos were recorded in short segments of no longer than six to eight minutes. This reduced editing during production and provided manageable segments of content for students to access and review. The tools used to produce and provide student access to these videos included a desk top visualiser to record hand calculations of worked examples, free screen capture software to record PowerPoint and Excel demonstrations, YouTube to upload the collection of short videos, and Blackboard to upload face to face lecture material and links to YouTube videos.

[Insert Table 2 here]

The intervention took place in Semester 1, 2013. To assess student performance, we compared quizzes three and four results in 2013 with results from other quizzes in 2013 and results from the 2011 and 2012 quizzes. The course content and course structure remained unchanged from 2011 to 2013. In the next section we expound our methodological approach, which allows us to control for student characteristics such as gender and other individual characteris-

tics, the delivery method and the final exam results. Section 3.2 contains a full description of the data used in this study.

2.1 Identifying student preferences for delivery methods

The Evaluation Unit at the University of Queensland’s Teaching and Educational Development Institute (TEDI) conducted an observation of student study habits while using videos and attending normal lectures. It was found that students generally liked using the online videos because it gave them more choice, control, accessibility and depth. However students missed the social interaction and motivation provided by the lecture and the opportunity to ask questions or hear the responses to questions asked by others. TEDI researchers also used a focus group and found students preferred to use a combination of face to face lectures with You Tube videos rather than to replace face to face lectures with only online lectures. Finally, an end of course survey (using Likert scale responses) was conducted and found that almost 90% of students did not want face to face lectures replaced with online versions of lectures. Students were equally divided on whether a blended learning approach should be implemented (using shorter lectures supplemented with YouTube videos) and almost 50% of students preferred the traditional two hour face to face lecture with no video lecture component.

The overwhelming conclusion from the observation and focus groups is that students have strongly manifested a preference for retaining live lectures. Yet there is a substantial cohort of students, approximately 30%, who would like to see the blended learning approach used more extensively. It is unclear, however, how much credence one should give to such manifested preferences as they may be subjected to the endowment effect – the notion that individuals assign more value to things simply because they own them.² Instead, our focus is on

²See, for example, Kahneman, Knetsch and Thaler (1990).

estimating the impact of blended learning on student performance, which we turn to in the next section.

3 Assessing the impact on student performance

A key methodological innovation used in this paper is the adoption of an experimental design that could be evaluated using the DID method. This allowed us to make two types of comparisons. First, we compared students' performance in 2013 when the experiment was implemented with the performance of students in 2011 and 2012. Second, as all students in the 2013 cohort participated in the experiment, we compared their performance with material delivered through blended learning and material delivered face to face. This method accounts for students' characteristics that might systematically influence their performance differently according to different delivery methods.

3.1 The difference-in-difference method

Consider the outcome of interest y_{ijt} (quiz scores) and the following linear regression model:

$$y_{ijt} = \alpha_0 + \alpha_1 \mathbf{X}_{it} + \alpha_2 \text{Blend}_{jt} + \psi_j + \mu_t + \epsilon_{ijt} \quad (1)$$

where i , j , and t are indices for a particular student, quiz number, and cohort respectively. $\mathbf{X}_{it}$ includes observed student's characteristics for cohort t , Blend_{jt} is an indicator of quiz j that assesses the topics for which lectures are delivered by blended learning for cohort t , and ψ_j and μ_t represent quiz-specific and cohort-specific effects respectively. Let $\epsilon_{ijt} = u_{it} + v_{jt} + \phi_{ijt}$ where u_{it} and v_{jt} are unobserved student-cohort and quiz-cohort specific characteristics and ϕ_{ijt} includes remaining unobserved factors. Estimation of the equation (1)

by the OLS provides an unbiased estimate of the parameter of interest α_2 , the effect of exposure to blended learning, only if

$$E(\epsilon_{ijt} | Blend_{jt}, \mathbf{X}_{it}, \psi_j, \mu_t) = E(u_{it} + v_{jt} + \phi_{ijt} | Blend_{jt}, \mathbf{X}_{it}, \psi_j, \mu_t) = 0. \quad (2)$$

This is only true when $cov(u_{it}, Blend_{jt} | \mathbf{X}_{it}, \psi_j, \mu_t) = 0$, $cov(v_{jt}, Blend_{jt} | \mathbf{X}_{it}, \psi_j, \mu_t) = 0$ and $cov(\phi_{ijt}, Blend_{jt} | \mathbf{X}_{it}, \psi_j, \mu_t) = 0$. Otherwise, omitted factors, ϵ_{ijt} , that are correlated with $Blend_{jt}$ can cause bias for α_2 . For instance, suppose that students with high motivation (or any other unobserved factors that could positively affect quiz scores) prefer face to face lecture. Then, negative correlation between $Blend_{jt}$ and motivation and positive effect of motivation on score could cause downward bias for α_2 . Below we argue that these three conditions are satisfied.

First, note that in general zero correlation is not guaranteed so that $cov(u_{it}, Blend_{jt} | \mathbf{X}_{it}, \psi_j, \mu_t) \neq 0$. That is, it might be that students who participated in the blended learning experiment were somehow different from students in other cohorts. For example, as in the case of many blended learning experiments described in the existing literature, students could select whether or not to participate in the experiment. This could lead to possible biases as students' choices might be correlated with unobserved factors that affected students' marks. However, our experiment design guarantees that $cov(u_{it}, Blend_{jt} | \mathbf{X}_{it}, \psi_j, \mu_t) = 0$. This is because once we account for cohort-specific effects (μ_t), all students in the same cohort should take the same quizzes so that students cannot self-select into or opt out from blended learning and face to face learning.

Second, we obtain $cov(v_{jt}, Blend_{jt} | \mathbf{X}_{it}, \psi_j, \mu_t) = 0$ because, in our experimental design, textbook, course content, difficulty of the quizzes and all other factors except lecture delivery mode remained unchanged over the three years

and we account for ψ_j . Finally, we obtain $cov(\phi_{ijt}, Blend_{jt} | \mathbf{X}_{it}, \psi_j, \mu_t) = 0$. This is because unobserved factors can vary in all i , j , and t simultaneously only if a particular student i in cohort t behaves differently (say students work harder/less for a particular quiz with $Blend_{jt} = 1$) across different quiz j . However, there is no reason for students to behave differently in this systematic way by exerting more effort on a particular quiz.

It follows then that the OLS estimation of the equation (1) that accounts for ψ_j and μ_t provides a causal estimate for α_2 . Note that we can interpret α_2 in the equation (1) as a DID estimator. To see this, consider the following example. Suppose there are only two periods, $t = 2013$ and 2012 . Consider over-time differencing of scores between quizzes that assessed material covered via blended learning (i.e. $y_{ijt}^{b,j^{34},2013} | Blend = 1, j = 3, 4, t = 2013$) and those that assessed material not covered by blended learning (i.e. $y_{ijt}^{f,j^{34},2012} | Blend = 0, j = 3, 4, t = 2012$). Defining $\Delta y_{ijt}^1 = y_{ijt}^{b,j^{34},2013} - y_{ijt}^{f,j^{34},2012}$ and applying (1) we obtain:

$$\Delta y_{ijt}^1 = \alpha_1 \Delta \mathbf{X}_{it} + \alpha_2 + \Delta \mu_t + \Delta \epsilon_{ijt} \text{ for } j = 3, 4 \text{ and } t = 2013. \quad (3)$$

Similarly, we can measure the over-time difference of scores from other remaining non-treated quizzes (that is, quizzes covering material not delivered by blended learning) between cohorts that were exposed to blended learning (i.e. $y_{ijt}^{f,j^{1256},2013} | Blend = 0, j = 1, 2, 5, 6, t = 2013$) and those that were not exposed to blended learning (i.e. $y_{ijt}^{f,j^{1256},2012} | Blend = 0, j = 1, 2, 5, 6, t = 2012$). Defining $\Delta y_{ijt}^0 = y_{ijt}^{f,j^{1256},2013} - y_{ijt}^{f,j^{1256},2012}$ and using (1) we obtain:

$$\Delta y_{ijt}^0 = \alpha_1 \Delta \mathbf{X}_{it} + \Delta \mu_t + \Delta \epsilon_{ijt} \text{ for } j = 1, 2, 5, 6 \text{ and } t = 2013. \quad (4)$$

Finally, using equations (3) and (4), our DID estimator then corresponds to

$$E(\Delta y_{ijt}^1 - \Delta y_{ikt}^0) = \alpha_2 \text{ for } j = 3, 4, k = 1, 2, 5, 6 \text{ and } t = 2013 \quad (5)$$

To be clear, (5) identifies the pure effect of exposure to blended learning in the sense that it removes any influence from unobserved quiz-specific and cohort-specific characteristics. That is, the differences (3) and (4) remove unobserved quiz-specific characteristics that are common across cohorts while the difference (5) removes unobserved cohort-specific characteristics that are common across quizzes.

Figure 1 graphically illustrates why the DID estimator represents an unbiased estimator of α_2 . Point A represents the mean score of quizzes assessing material covered with blended learning (quizzes three and four) in untreated years (2011/2012). Point B represents the mean score of quizzes assessing material covered by face to face learning (quizzes one, two, five and six) in untreated years. The distance between A and B (or between C and D) represents the inherent average difference in difficulty between quizzes 3,4 and quizzes 1,2,5,6. This quiz-specific effect is accounted for by the differences (3) and (4) which remove ψ_j . Point E represents the mean score of quizzes with blended learning in the treated year (2013) and point D represents the counter-factual mean score of quizzes 3,4 in treated year if blended learning was not provided in 2013. Finally, point C represents the mean score of quizzes assessing material covered by face to face learning in the treated year.

It follows that the distance between A* and D (and between B* and C) represents the average cohort score difference between the two years. This cohort-specific effect is accounted for by the second differencing in (4) which removes μ_t . Therefore, equation (3) becomes $E-A=E-A^*$, equation (4) becomes $C-B=C-B^*$. Finally, equation (5) becomes $E-A^*-(C-B^*)$ and this is the negative of the distance between D and E. This is the DID effect we es-

timate by α_2 in the regression equation (1). A critical assumption for the validity of the DID estimate is the assumption of the same slope for BC and AD and this is guaranteed by $E(u_{it} - u_{it-1})|Blend_{jt}, \mathbf{X}_{it}, \psi_j, \mu_t) = 0$ and $E(v_{jt} - v_{jt-1})|Blend_{jt}, \mathbf{X}_{it}, \psi_j, \mu_t) = 0$ from our experimental design, which in general could be easily violated in other experimental settings mainly due to self-selection and heterogeneity of treated and untreated assessments.

3.2 Data

The data consists of 2,071 observations on students' characteristics including age, gender, international versus domestic status, language spoken at home, single versus double degree status, final exam and quiz scores. The data was de-identified before being sent to us. Summary statistics are provided in Table 3 below. It should be noted that the decrease in the number of students undertaking the course in Semester 1, 2013 can be accounted for by its removal as a compulsory course in the Bachelor of Business Management degree from 2013 onwards. This degree is very popular amongst international students as illustrated by the decrease in the percentage of international students taking the class in 2013 vis-a-vis 2011/2012.

[Insert Table 3 here]

3.3 Empirical Results

Table 4 reports results of OLS and DID estimations.³ The OLS results in columns (1) and (2) indicate statistically significant negative effects of blended learning. Blended learning reduced quiz scores by 3.4 points in a hundred point scale (14 percent of one standard deviation) and 1.9 points (8 percent of one

³We explored various specifications that allow nonlinearity by taking log and squares of variables and interactions among included variables and we obtain quite robust results. We only report the results from the simplest specification to save space.

standard deviation) without and with controlling for students' characteristics respectively. Including extensive control variables mitigates negative effect of blended learning but it is still statistically significant. However, the OLS estimation cannot control for unobserved cohort-specific and quiz specific effects that could be correlated with blended learning status.

The DID results are reported in columns (3) and (4) without and with controlling for students' characteristics respectively and the results show that blended learning has no effect on quiz score. The estimated DID effect of blended learning is the increase of score by 0.16 (0.6 percent of one standard deviation) points and its p-value is 0.86 without control variables and the decrease of score by 0.18 (0.7 percent of one standard deviation) points and its p-value is 0.83 with control variables. This implies that the OLS estimation cannot completely eliminate bias due to self-selection and heterogeneity of treated and untreated assessments even with extensive students' characteristics. This is strong evidence that student performance is not affected by the introduction of blended learning.

[Insert Table 4 here]

Several previous studies provide evidence that the effect of online learning could vary across gender (Ferber 1995; Brown and Liedholm 2002; Figlio et al. 2013), race (Brown and Liedholm 2002; Figlio et al. 2013), academic achievement level (Figlio et al. 2013) among others. We explore the differential effects of blended learning across students' characteristics (we denote these variables as $\mathbf{z}$) such as age, gender, nationality (both birth and current), primary language, and academic achievement level.

Table 5 reports results for the DID estimates of differential effects of blended learning. To save space, we only report the DID estimation results with the specification that all control variables and both cohort and quiz fixed effects

are included. Each column reports the results of two regressions with two subsamples. For each column, the first row reports the effect of blended learning with the sample restricted to $z = 0$ and the other row reports the effect with the sample of $z = 1$. Except for column (2) which reports differential effects across gender, blended learning has no statistically significant effects on student performance. In other words, student performance is not affected by the introduction of blended learning regardless of students' characteristics such as age, nationality, primary language, and achievement level. Interestingly, the effect of blended learning is negative for male students while it is positive for female students.

This result is consistent with Shea et al. (2001) and Brown and Liedholm (2002) that female students perform more favorably with online learning. Female students gained 2.4 points in blended learning compared to face to face learning while male students lost 2.3 points. Although the effects are statistically significant, the magnitude is very small as they are less than 10 percent of one standard deviation.

[Insert Table 5 here]

4 Discussion and concluding remarks

This paper reports the outcome of an experiment with blended learning at the University of Queensland's School of Economics introductory statistics course. The experiment involved the introduction of blended learning for two out of the thirteen lectures.

The experimental design allowed us to test the impact of blended learning on student performance by utilising a differences in differences approach. In particular, this approach takes into account any possible inherent differences in unobserved characteristics of the students who were exposed to blended learning

vis-a-vis those who were not and differences in the level of difficulty of the quizzes.

Our results strongly suggest that blended learning has no impact on student performance and the results are very robust to various specifications that allow nonlinearity and interactions of variables. Moreover, student performance is not affected by the introduction of blended learning regardless of students' characteristics such as age, nationality, primary language, and achievement level. Consistent with the existing literature, the effect of blended learning is negative for male students while it is positive for female students. This differential impact of blended learning on gender is yet to be explained.

As indicated previously, there are many benefits associated with the introduction of blended learning. These include, but are not limited to, an ability to engage with students using media that they are accustomed to (e.g., YouTube) and reduced costs of delivering material online in comparison to face to face delivery. This paper provides strong evidence that blended learning has no detrimental effect on student performance.

References

- [1] Allen, I. E. and Seaman, J. (2006). Making the grade: Online education in the United States. Boston: Sloan Consortium.
- [2] Al-Qahtani, A. and Higgins, S. E. (2013). Effects of traditional, blended and e-learning on students' achievement in higher education. *Journal of Computer Assisted Learning*, 29, pp. 220-234.
- [3] Brown, B. W., and Liedholm, C. E (2002). Can Web courses replace the classroom in principles of microeconomics? *American Economic Review* 92(2), pp. 444-448
- [4] Donovan, C., Figlio, D., and Rush, M. (2006). Cramming: The effects of school accountability on college-bound students. Working Paper, National Bureau of Economic Research, Cambridge, MA. .
- [5] Dutton, J., Dutton, M., and Perry, J. (2001). Do Online Students Perform as Well as Lecture Students? *Journal of Engineering Education*, 90(1), pp. 131-136.
- [6] Exetera, D. J. , Ameratungaa, S. , Ratimab, M. , Mortona, S., Dicksona, M, Hsua, D. and Jackson, R. (2010). Student engagement in very large classes: the teachers' perspective *Studies in Higher Education* 35(7), 761–775.
- [7] Ferber, M. A. (1995). The Study of Economics: A Feminist Critique. *American Economic Review* 85(2), pp. 357-361
- [8] Ferguson, J. and Tryjankowski (2009). Online versus face to face learning: looking at modes of instruction in Master's level courses, *Journal of Further and Higher Education*, 33(3), pp. 219-228.

- [9] Figlio, D. , Rush, M. and Yin, L. (2013). Is it live or is it internet? Experimental estimates for the effects of online instruction on student learning. *Journal of Labor Economics* 31(4), pp. 763-784.
- [10] Johnson, H. D., Dasgupta, N., Zhang, H. and Evans, M. A. (2009). Internet Approach versus Lecture and Lab-Based Approach for Teaching an Introductory Statistical Methods Course: Students' Opinions. *Teaching Statistics* 31(1), pp. 21-26.
- [11] Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1990). Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy* 98(6), pp. 1325-1348.
- [12] Kartha, C. P. (2006). Learning business Statistics: Online Vs Traditional. *The Business Review, Cambridge*, 5(1), pp. 27-32.
- [13] McLaren, C. H. (2004). A comparison of student persistence and performance in online and classroom business statistics experiences, *Decision Sciences Journal of Innovative Education*, 2(1), pp 1-10.
- [14] Mortensen, T. (2005). State tax fund appropriates for higher education: FY1961 to FY2005. Report, Post secondary Education Opportunity, 151, pp 1-7.
- [15] Morris, D. (2008). Economies of scale and scope in e-learning. *Studies in Higher Education*, 33(3), pp. 331-343.
- [16] Sebastianelli, R. and Tamimi N. (2011). Business Statistics and Management Science Online: Teaching Strategies and Assessment of Student Learning, *Journal of Education for Business*, 86(6), pp. 317-325.
- [17] Shea, P., Fredericksen, E., Pickett, A., Pelz, W., and Swan, K. (2001). Measures of learning effectiveness in the SUNY learning network, in Bourne,

- J. and Moore, J. C., eds., *Online education, Volume 2. Learning effectiveness, faculty satisfaction, and cost effectiveness*. Needham, MA: Sloan Center Online Education, pp. 31-54
- [18] Suanpang, P., Petocz, P., and Reid, A. (2004). Relationship between learning outcomes and online accesses. *Australasian Journal of Educational Technology*, 20(3), pp. 371-387.
- [19] Twigg, C. A. (2013). Improving learning and reducing costs: Outcomes from changing the equation. *Change: The Magazine of Higher Learning* 45(4), pp. 6-14.
- [20] US Department of Education (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Washington, DC.

Table 1: Selected references on assessment of online and blended learning

Methodology	Choice FtoF vs OL/BL	OL/BL replaces FtoF
Survey/focus groups	Katha(2006) ⁼	Suanpang et al.(2004) ⁻ Johnson et al.(2009) ⁺
Assessing grades	Ferguson&Tryjankowski(2002) ⁺ Dutton,Dutton&Perry(2001) ⁻ Brown&Liedholm(2002) ⁺	Al-Qahtani&Higgins(2013) ⁼ Figlio et al.(2013) ⁼ McLaren(2004) ⁼

Notes: "=" means no difference between the two approaches; "+" means the FtoF does better than either OL or BL and "-" means either OL or BL does better than FtoF

Table 2: Treatment and control

	quiz1	quiz2	quiz3	quiz4	quiz5	quiz6
2011	FtoF	FtoF	FtoF	FtoF	FtoF	FtoF
2012	FtoF	FtoF	FtoF	FtoF	FtoF	FtoF
2013	FtoF	FtoF	BL	BL	FtoF	FtoF

Note: "FtoF" represents face to face learning and "BL" represents blended learning

Table 3: Basic Statistics: Means and Standard Deviations

	(1)	(2)	(3)
variable	Overall	2011/2012	2013
<i>age</i>	19.37 (2.13)	19.50 (2.12)	19.02 (2.14)
<i>female</i> (= 1)	0.45 (0.50)	0.46 (0.50)	0.44 (0.59)
<i>international</i> (= 1)	0.25 (0.43)	0.27 (0.44)	0.19 (0.39)
<i>English</i> (= 1)	0.68 (0.46)	0.68 (0.47)	0.71 (0.46)
<i>dual degree</i> (= 1)	0.44 (0.49)	0.41 (0.49)	0.51 (0.50)
<i>finalexam</i>	68.55 (19.47)	69.61 (19.29)	65.76 (19.68)
quiz 1	82.65 (15.21)	81.73 (15.57)	85.01 (13.98)
quiz 2	71.57 (19.39)	70.44 (20.16)	74.36 (17.03)
quiz 3	67.34 (24.40)	67.26 (24.54)	67.56 (24.54)
quiz 4	70.70 (25.08)	71.13 (24.87)	69.59 (25.61)
quiz 5	73.79 (20.40)	73.94 (20.47)	73.41 (20.23)
quiz 6	66.11 (23.33)	66.45 (23.63)	65.23 (22.53)
No. of obs.	2,071	1,490	581

Notes: Scores for quizzes and final exam are scaled to have values from 0 to 100. Standard deviations of variables are reported in parentheses.

Table 4: The effects of face-to-face learning versus blended learning

	(1)	(2)	(3)	(4)
	OLS		DID	
<i>Blended</i>	-3.42*** (0.90)	-1.86** (0.74)	0.16 (0.89)	-0.18 (0.85)
<i>age</i>		-0.28** (0.16)		-0.22 (0.15)
<i>female</i>		1.80*** (0.55)		1.86*** (0.55)
<i>international</i>		0.06 (0.95)		0.29 (0.94)
<i>English</i>		-2.48*** (0.78)		-2.37*** (0.77)
<i>final exam</i>		0.56*** (0.02)		0.59*** (0.01)
FEs	No	No	Yes	Yes
Control variables	No	Yes	No	Yes
R^2	0.01	0.27	0.07	0.34
No. of obs.	11,348	11,245	11,348	11,245

Notes: FEs represents quiz-specific and cohort-specific fixed effects. Control variables include student characteristics age, final exam score, and the dummy variables such as gender, nationality, primary language(English), dual degree, major, and birth place. Cluster (student) robust standard errors are reported in parentheses. ***, ** and * indicate that the coefficient is statistically significant at 1%, 5%, 10% level, respectively.

Table 5: Heterogenous effects of face-to-face learning versus blended learning

	(1)	(2)	(3)	(4)	(5)
	DID				
<i>Blended</i> $z = 0$	0.37 (1.01)	-2.28* (1.19)	-0.24 (0.98)	0.48 (1.46)	-0.28 (1.62)
<i>Blended</i> $1(age > 19)$	-1.50 (1.58)				
<i>Blended</i> <i>female</i>		2.41** (1.19)			
<i>Blended</i> <i>international</i>			1.05 (1.68)		
<i>Blended</i> <i>English</i>				-0.48 (1.04)	
<i>Blended</i> <i>highachievers</i>					0.87 (0.94)
FEs	Yes	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes

Notes: z includes dummy variables for $1(age > 19)$, female, international, English, and high achievers. FEs represents quiz-specific and cohort-specific fixed effects. Control variables include student characteristics age, final exam score, and the dummy variables such as gender, nationality, primary language(English), dual degree, major, and birth place. Cluster (student) robust standard errors are reported in parentheses. ***, ** and * indicate that the coefficient is statistically significant at 1%, 5%, 10% level, respectively.

Figure 1: Illustration of the DID effect of blended learning

