Student Performance in an Introductory Business Statistics Course: Does Delivery Mode Matter?

Jonathan Haughton **Alison Kelly**

jhaughton@suffolk.edu akelly@suffolk.edu

Department of Economics

Suffolk University, Boston MA 02108

Abstract – Approximately 600 undergraduates completed an introductory business statistics course in 2013 in one of two learning environments at Suffolk University, a mid-sized private university in Boston, Massachusetts. The comparison group completed the course in a traditional classroom-based environment, whereas the treatment group completed the course in a flipped-hybrid environment, viewing lecture material online prior to once-a-week face-to-face meetings. After controlling for observable differences, students in the hybrid environment performed better on the common final exam; however, there were no significant differences in the final grades or student satisfaction between the two environments.

Keywords: hybrid, student performance, business statistics, selection bias

Acknowledgements: We would like to thank Jim McCarthy for his most helpful comments, Zi (Bennie) Yang for carefully reading the document, Katie Linder for her suggestions, and participants at the economics department seminar at Suffolk University for useful feedback.

Concerns about access and quality are among the perennial challenges facing higher education. Greater use of technology in teaching is widely seen as a promising way of controlling costs (and hence potentially improving access), and reducing achievement gaps. Yet, as online and hybrid courses proliferate across college and university campuses, there has been much controversy about the effectiveness of these newer modes of instruction, relative to that of a traditional face-to-face (F2F) format.

In this paper, we evaluate whether a flipped hybrid (or “blended”) model for teaching introductory business statistics performs better or worse than the more traditional class-intensive face-to-face approach to teaching. By “flipped hybrid” we mean a model of course delivery where students are expected first to study short on-line videos that present the material, test their understanding with on-line questions, and then come to a weekly face-to-face problem-solving class with a professor. The data come from Suffolk University in Boston, a typical medium-sized university where half of all undergraduates take STATS250 (Applied Statistics), and where half of the sections of the course are now delivered as flipped hybrids. Given the importance of courses in introductory business statistics in most universities, the representative nature of the Suffolk undergraduate body, the relatively large sample size, and our ability to control for many potential influences on performance, we believe that our results are broadly applicable.

Why Consider Flipped Hybrids?

The traditional model of teaching introductory business statistics has been remarkably durable. Typically, students have two or three F2F classes, for a total of 2½-3 hours, every week. The teacher presents the material, assigns homework exercises, and tests students with quizzes and exams.

A flipped hybrid course departs from the traditional model in two important ways. First, it reverses the sequence; students are first introduced to the substantive material on-line, and then come to class where the teacher helps clarify points of confusion or difficulty, and adds anecdotes, examples and extensions to the basic analysis. Second, it puts more emphasis on self-directed on-line activity, and less on spending time in the classroom; typically, there is just a single F2F class per week, with half as much face time with a professor as in the traditional model.

The flipped hybrid model has a number of potential attractions. It may be a very effective approach to teaching introductory business statistics, which has a clear body of concepts and techniques that need to be understood, mastered, and applied. Even if hybrids do not work for everyone, they might be well-suited to segments of the student body – for instance, to students who grasp the material quickly, and do not need to spend as much time in F2F classes, or to students who need extra time to master the material and for whom one-size-fits-all classes are too short. This would address one of the biggest challenges faced by teachers of statistics, which is the wide dispersion of student aptitude for statistics; the final grade distribution is often bimodal.

As compared to a course offered entirely online, the professor-student relationship within the classroom remains intact with the hybrid model. When Sebastianelli and Tamimi (2011) assess the quality of quantitative business courses offered entirely online, they find the features involving professor-student interaction to be the most useful. Moreover, in terms of learning the quantitative content, they conclude that discussion forums are of limited value, and features involving student-student interaction are the least useful.

The flipped hybrid model may also be cheaper; by halving the amount of F2F time, the cost of faculty and classrooms can potentially be halved. On the other hand, there may be substantial up-front costs in creating suitable on-line materials; faculty may not be willing to teach twice as many students, even if contact hours remain unchanged, given the potentially increased traffic during office hours, a heavier burden of grading, and less-satisfying relations with students whom one meets only once a week.

Three other considerations are relevant. Changes in technology have made on-line materials cheaper, better, and more accessible, so the quality of the on-line part of a hybrid course has improved compared to even a decade ago. Students, raised in a more virtual world, may be more receptive to on-line learning than their parents were, and more faculty members may be comfortable teaching in this way. And whether we like it or not, the flipped hybrid course has become fashionable; before rushing to embrace this pedagogy, we need to evaluate whether it delivers on its promise.

One of the first studies to gain widespread attention on the effectiveness of online instruction was released in 2009 (and updated in 2010) by the U.S. Department of Education (Means et al. 2010). This study, a meta-analysis of the then-available research on online learning, found that online courses were more effective at satisfying learning outcomes as compared to F2F courses, with a hybrid format having the largest benefits of all. However, the methodology used by the DOE study has been criticized on a number of fronts: most importantly, none of the studies in the meta-analysis included randomly-assigned students taking a full-term course in settings that could be directly compared (i.e., similar instructional materials by the same instructor, or a standardized course). Moreover, only seven of the 45 studies in the DOE meta-analysis – already chosen from 1,132 studies published between 1996 and 2008 on the basis of their rigor – involved undergraduate or graduate students enrolled in semester-long courses, and these seven studies found no significant differences in outcomes between online and F2F formats. A more recent survey by Lack (2013) identifies about 30 subsequent relevant and acceptably rigorous studies, and concludes that "the literature … yields little, if any, evidence to suggest that online or hybrid learning, on average, is more or less effective than face-to-face learning" (p.10).

As noted by Lack (2013), few studies control for pre-existing effects when measuring the impact of online courses. Lam (2009) used regression analysis to assess student performance in traditional and online formats of an undergraduate computer programming course. She finds that delivery mode had no significant effect on student performance, and that cumulative GPA was the only significant predictor. Ary and Brune (2011) compared learning outcomes in traditional and online formats of a personal finance course. Their regression results suggest that the delivery mode did not significantly influence course averages, but the percentage change in scores between pre- and post-tests was significantly higher for the traditional format.

Wilson and Allen (2011) assessed the success rates of F2F and online students in two different business courses. They find that withdrawal rates and failure rates were not significantly different between the two modes of course delivery. They too conclude that cumulative GPA was the greatest predictor of course grade, regardless of delivery mode. Driscoll et al. (2012) compared student performance and satisfaction between F2F and online sections of an introductory sociology course that was taught by one instructor over multiple terms with very little change in course materials and assessments. They find no significant difference in student performance or student satisfaction between the two different formats. The papers by Lam, Ary and Brune, Wilson and Allen, and Driscoll et al., all include controls for a number of background characteristics and/or other predictor variables; however, these studies are quasi-experimental in that they do not assign students randomly to F2F or online courses. Randomization, which would be required to avoid self-selection bias, is difficult to implement on college campuses for a variety of reasons, including Institutional Review Board requirements, some students' reluctance to comply with their random assignments, and logistical issues related to the scheduling of classes.

Figlio et al. (2010) conducted an experiment in which students were randomly assigned to either a F2F or online section of a Principles of Microeconomics course taught by one instructor. The only difference between these sections was the mode of delivery: students either attended live lectures, or watched these same lectures in an Internet setting. All other ancillaries for the class, such as problem sets and exams, were the same. A simple means comparison over three exams shows that students performed better in the live setting as opposed to the online setting, but the differences were not statistically significant. When Figlio et al. control for student race/ethnicity, sex, and prior achievement levels, they find that test scores are significantly higher for Hispanic students, male students, and low-achieving students in the case of live lectures.

Bowen et al. (2012) conducted an ambitious semi-randomized study assigning 605 students on six public college campuses to take either a traditional or hybrid version of an introductory statistics class.[1] The hybrid course used a prototype machine-guided mode of instruction developed at Carnegie Mellon University in concert with one F2F meeting each week. The authors find no statistically significant differences in learning outcomes between students in the traditional- and hybrid-format sections.

The Relevance of STATS250

The results of the Bowen et al. study are important and useful, but we are interested in knowing whether they still apply when the on-line instruction does not use the Carnegie Mellon prototype, when hybrid courses are expanded to a larger proportion of the class, and when hybrid courses become a banal part of the curriculum and the initial novelty wears off. The case of the introductory business statistics course at Suffolk University (STATS250) allows us to address these issues, and thus is relevant for a wide swathe of university-level business statistics courses in the United States.

Every semester, between 250 and 300 students, three-fifths of them sophomores, enroll in STATS250. The 4-credit course is required of students in business-related disciplines and economics, but attracts significant numbers of students from other fields, and over half of all undergraduates take the course at some point.[2] Up and until the fall semester of 2012, STATS250 was taught in a traditional format, with about 10-12 sections per semester capped at 30 students each. Two flipped hybrid sections were introduced in the spring of 2013, alongside eight traditional sections; by fall 2013, half of the twelve sections were hybrids. The introduction of flipped hybrid sections provides an opportunity to evaluate their impact, especially as students in all sections – traditional and hybrid – take a common exam at the end of the semester. By summer 2014, over 350 students will have taken the hybrid version of STATS250, a larger group than even the sample studied by Bowen et al. (2012).

Before the start of a semester, students choose to enroll in either a traditional or hybrid section of STATS250. Students who register for the traditional format may choose a section that meets once a week for two hours and 40 minutes, twice a week for 75 minutes, or three times a week for 50 minutes. Students enrolled in the hybrid format meet once a week for 75 minutes with the instructor. Prior to each weekly meeting, students in all hybrid sections are required to complete the same assigned textbook reading, view some on-line video clips, and complete a set of conceptual online exercises. These on-line materials are part of LearnSmart, a component of McGraw-Hill's Connect product, and accompany the course textbook written by Jaggia and Kelly (2012). At the weekly face-to-face meeting, students complete the same set of in-class exercises and case studies, and discuss any difficulties with the teacher. Within five days of the class meeting, students are required to submit the same online homework assignments (which generate questions randomly from a test bank). In both the hybrid and the traditional sections, all quizzes and exams are administered in a classroom environment.

The introductory business statistics course at Suffolk University is a relevant model for many other universities. The material covered by the course is standard, the course uses a textbook from McGraw-Hill, a major publisher, and the level of student preparation is similar to that of hundreds of other institutions. The Carnegie Foundation classifies Suffolk University's undergraduate instructional program as having "balanced arts and sciences/some graduate coexistence", a category that includes 291 universities with 2 million students, or 10% of the national student body (Carnegie Foundation 2014). A private, coeducational, non-sectarian university located in downtown Boston, Suffolk has about 8,800 students, including 5,800 undergraduates; it is thus slightly larger than the average four-year university in the United States, which has 4,600 students (NCES 2014, Tables 301.10 and 317.10). Suffolk University's Sawyer School of Business is AACSB accredited. The average SAT score, for reading and mathematics, of entering undergraduates is 1,050, and the university is considered to be "selective" by U.S. News (2014). In addition, in its Regional University North Rankings, Suffolk University is ranked 60 out of 135 universities.[3] The relevance of our study to the teaching of introductory business statistics elsewhere comes from the standard nature of the course content of STATS250, the large sample size, and the representativeness of the students taking the course and the institution in which they are enrolled.

Research Design

We use four distinct outcome measures in our examination of the impact of flipped hybrid classes. The first is the most straightforward: at the end of every semester, every student in STATS250 takes a common final exam, which allows one to compare the performance of students in hybrid sections with that of students in traditional F2F sections. The final exams used at the end of fall 2012 and spring 2013 were identical, and so are comparable; subsequent final exams differed somewhat from one semester to the next.

The second measure of performance is the letter grade on the course, which varies from fail (0 on a 4-point scale) to A (4 on the scale). The semester grade is based on assignments, quizzes, and midterms in addition to the common final exam. Different teachers may determine letter grades differently, and this lack of consistency across sections makes this a less compelling outcome measure.

The two other measures of performance are subjective. One is based on student responses to a question that asks, “Overall, how would you rate STATS250 relative to your other courses?”, and records the answers on a scale of 1 (much worse) through 5 (much better). The other is based on student answers to a question that asks, “How much did STATS250 increase your interest in the subject matter?”, and again records the answers on a five-point scale that runs from 1 (not at all) through 5 (a great deal). All four measures are designed so that larger numbers are associated with better outcomes.

The data for the impact evaluation come from three sources:

1. The grades for the final exam, which were compiled by the course coordinator;
2. Information on student attitudes, which was collected using a questionnaire that students were asked to complete immediately after they finished the final exam. Students were asked to provide their ID numbers, but not name, gender, age, or other distinguishing personal identifier. The questionnaire, which may be found via our web site (http://web.cas.suffolk.edu/faculty/jhaughton/), is modeled on the one used by Bowen et al. (2012), suitably adapted to the context of Suffolk University.
3. Final grades and background information on students, including their cumulative GPA, course load, and admissions ranking, which were obtained from the Registrar’s office.

Permission from the University’s Institutional Review Board was requested, and granted, for this research, given that it involves human subjects.

Data from the three sources of information were matched using student ID numbers. Unfortunately, 26 of those who completed the student survey in spring 2013 (and 25 in fall 2013) did not report their ID number, or reported a number that could not be matched elsewhere. A further nine students in spring 2013 (and 42 in fall 2013) did not complete the questionnaire, and in each semester 11 students did not take the final exam. Thus, of the 279 students listed on the final exam roster in the spring (and 309 in the fall), there is missing data for 46 (16.5%) in the spring and 78 (25.2%) in the fall, as Table 1 shows. This raises the possibility that response bias may be a problem. The lower panel in Table 1 addresses this issue, by comparing known information from responders and “non-responders” (i.e. those who did not take the final exam, or fill in a questionnaire, or give a usable ID on a completed questionnaire). The non-responders performed less well academically than responders, but in most other respects look very similar to responders, suggesting that response bias is not likely to be a serious problem.

Baseline summary statistics on the four outcomes are shown in Table 2. There were no statistically significant differences in grades between fall 2012 and spring 2013, either in the common final exam (67.1% vs. 68.1%) or the semester grade (2.6 vs. 2.5, or just under a B-). If the student responses to the course rating, and to the extent to which the course raised their interest in statistics, are converted to a continuous scale, then there was no discernible difference between fall 2012 and spring 2013. It is interesting that the average course grade was, at about C+, significantly lower in fall 2013 than in previous semesters; the rating of the course, and the interest in statistics that the course engendered, also fell significantly in fall 2013.

We now examine the average outcomes of hybrid classes, compared to traditional ("non-hybrid") classes; the relevant numbers are set out in Table 3. Both in spring 2013 (when two of 10 classes were hybrids) and fall 2013 (when six of 12 classes were hybrids), students in the hybrid classes obtained higher scores on the common final exam, and this difference is statistically significant at the 10% level. However, course grades were not significantly higher for hybrid sections, and although students in hybrid courses enjoyed statistics, and were turned on by the subject, in spring 2013, there was no such effect in fall 2013.

The simple comparisons in Table 3, although suggestive, have a serious flaw: they do not take into account the possibility of selection bias. Students were not randomly assigned to hybrid courses, and it is entirely possible that students who chose to enroll in the hybrid sections were atypical – perhaps they were more self-motivated, or academically stronger, or systematically different from their peers in some relevant way.

Addressing Selection Bias

There are a number of ways to address selection bias, although in the absence of panel data, none are able to deal satisfactorily with the unobservable characteristics that might impel a student to enroll in a hybrid rather than a traditional section of the course (Haughton and Haughton 2011, chapter 12). If we are willing to assume that, after controlling for observable characteristics of the student, he or she ended up taking the hybrid course randomly, then we have partial randomization, and may apply the techniques of quasi-experimental design.

The most straightforward of these quasi-experimental methods is to estimate a regression where the dependent variable is one of the relevant outcomes, and the "treatment" – i.e. whether a student is enrolled in a hybrid course – is included as a dummy variable on the right hand side. Suppose that we may assume, for the treated cases, that the outcome Y_i depends on control variables X_i as follows:

$$Y_i^T = \alpha^T + X_i\beta^T + v_i^T, i = 1, \dots, n_1 \tag{1}$$

and, for the non-treated ("comparison" or "control") cases,

$$Y_i^C = \alpha^C + X_i\beta^C + v_i^C, i = 1, \dots, n_2 \tag{2}$$

where the error terms are assumed to be normally distributed with zero means and constant variances. Pooling the data for the treatment and comparison samples we get the switching regression:

$$Y_i = \alpha^C + (\alpha^T - \alpha^C)T_i + X_i\beta^C + X_i(\beta^T - \beta^C)T_i + \varepsilon_i \ \ , i = 1, \dots, (n_1 + n_2) \tag{3}$$

where T_i is set to 1 if the student is treated and to 0 otherwise. This reduces to the common impact model if we assume, as is often done, that $\beta^T = \beta^C$:

$$Y_i = \alpha^C + (\alpha^T - \alpha^C)T_i + X_i\beta^C + \varepsilon_i \ \ , i = 1, \dots, (n_1 + n_2) \tag{4}$$

Our interest is in estimating the coefficient $(\alpha^T - \alpha^C)$. Estimates of this common impact model are reported in Table 4, for both spring and fall 2013. The dependent variable here is the percentage score on the common final exam, and the mean values of the variables are also shown, in order to provide a point of reference.

The regression estimates in Table 4 show clearly that in spring 2013, students in the hybrid sections obtained higher scores on the common final exam, after controlling for a large number of other variables. Indeed the magnitude of this effect, which is statistically highly significant, is slightly larger than the one found in the simple comparison in Table 3. However, the effect was no longer statistically significant in fall 2013.

A number of other features of the regression results in Table 4 – which we refer to as the "large model" because there are fully 35 independent variables – are noteworthy. Students who expected the course to be hard did relatively poorly, as did those who undertook more paid outside work. On the other hand, Chinese speakers, honors students, and those with a higher cumulative GPA, performed better. Compared to general business majors, students who majored in accounting, sciences, and perhaps economics, did better at the common final exam.

The lack of consistency in the estimated coefficients between the spring and fall versions of these regressions is striking, and may be due in part to underlying multicollinearity. One common response to the curse of dimensionality is to trim the model, using forward or backward stepwise regression; the key results are included in Table 6, and although the test statistics shown here are no longer strictly legitimate, the results for the treatment variable – which is our principal interest – are similar to those found with the large model. Thus, multicollinearity is not coloring the measure of the impact of the hybrid classes on outcomes.

The regression estimates do not necessarily eliminate selection bias (Ettner, undated); unobserved factors may contribute to a correlation between the error (ε_i) in equation (4) and the treatment dummy variable (T_i), leading to a biased estimate of the treatment effect ($\alpha^T - \alpha^C$). One practical solution is to use a matching technique; another is to estimate a treatment effects regression, which is effectively a form of Instrumental Variables regression.

A popular matching technique is propensity score matching (see for instance Haughton and Haughton 2011; Rosenbaum and Rubin 1983). First, one estimates a probit equation where the binary dependent variable is set to 1 if the student is enrolled in a hybrid section, and the independent variables reflect pre-existing conditions; the predicted values from this equation are the propensity scores. With nearest-neighbor matching, the next step is to match each treated student with the non-treated student who has the closest propensity score. The mean difference in outcomes between the treated students and their matches measures the average treatment effect on the treated (ATT). Propensity score matching has two strengths relative to the regression approach: it confines the comparison to observations in the region of "common support", which helps reduce selection bias, and it does not rely on distributional or parametric assumptions.

The estimates of the propensity score equations for spring and fall 2013 are shown in Table 5. We see that students who live independently were more likely to enroll in a hybrid course, while Spanish speakers were less likely to do so. Students in most business-related majors were also more likely to favor a hybrid section. The actual key results of the propensity score matching are given in Table 6, and show that in spring 2013, students in hybrid courses performed better on the common final exam (the difference was 9.05, p-value of 0.043), but did not differ from their peers on any of the other measures of outcome in the spring semester, or by any measure in the fall.

A related technique for measuring impact is direct matching; each student in a hybrid course is matched with a student in a traditional course by minimizing the Mahalanobis distance between them – based on the set of variables used in Table 5. Here too we find that in spring 2013, students in hybrid courses got higher scores on the common final exam, but their other outcomes were similar to those of students in non-traditional courses; and again there were no significant differences between hybrid and traditional classes, by any outcome measure, in fall 2013. The details are shown in Table 6.

We also used a treatment regression approach to measuring the impact of hybrid classes. Selection (or endogeneity) bias will occur if the error term in the treatment equation (Eqn. 4) is correlated with the treatment variable. One solution is to estimate a first-stage equation where treatment (T_i) is the dependent variable, and then to use the estimated (rather than actual) values of T_i from the first-stage estimation in the treatment equation. The use of predicted, rather than actual, T_i breaks the correlation between the residual and the treatment variable (Ettner, undated). The procedure works best if there are variables that are expected to affect whether one is treated (here, in a hybrid course) but do not affect the outcomes (such as final exam grades). We identified three such variables for spring 2013: whether a student is living at home, whether the student expressed a strong desire to learn statistics, and gender.

The relevant results of the treatment regression model for spring 2013 are shown in Table 6. By this measure, final exam scores for hybrid classes were substantially higher than for traditional classes in spring 2013 (but not in the fall). However, the value of λ, which is defined as the product of the correlation of the residuals of the two equations times the standard error of the outcome regression, was not significantly different from zero, suggesting that a single-equation method would have served adequately. The treatment regression results show no effect of hybrid classes on semester grade or the perceived course rating, but do indicate that students in the hybrid sections were more likely to show a greater interest in statistics after having taken the course (in spring 2013).

In the regression and matching exercises it is assumed (implicitly) that the introduction of hybrid sections does not affect the performance of traditional sections, but this assumption may be wrong. The use of hybrid sections economizes on the use of teachers – six teachers taught a total of 10 sections in fall 2012 and again in spring 2013, but only five teachers were needed for the 12 sections taught in fall 2013, half of which were hybrids. If teachers are hired in order of merit, the use of hybrid courses could allow one to avoid hiring the weakest instructors, thereby raising the average quality of instruction both for hybrid and traditional courses.

It is difficult to measure such an effect because it is rarely possible to construct a good counterfactual, since we do not usually know who might have been hired if there were fewer hybrid and more traditional sections. However the following thought experiment is useful: in fall 2013 there were six hybrid sections of STATS250, and if these were taught as traditional sections, then three additional instructors (each teaching one section) would have been required. Most of the adjunct teachers of statistics at Suffolk are students in the PhD program in economics, and we are able to identify who would likely have been hired. These three graduate students had taught STATS250 before in fall 2012 and/or spring 2013, and we know how their students performed on the final exam and rated the course, for those semesters. Applying these relative outcomes, we are able to simulate the effect on overall outcomes in fall 2013: the mean score on the final exam would have fallen from 66.6% to 65.9%, due in part to the lower performance of the students taught by the marginal hires (64.8%), and in part to the lower weight on the scores of the high-performing teachers (who would now teach fewer students). The difference is not statistically significant, however, and the other measures of performance (course grade, student evaluation of the course, and changed interest in statistics) would barely change. This exercise is hardly conclusive, but does suggest that the pool of effective statistics teachers may often be sufficiently deep that the indirect effect of hybrid courses, via an increase in the average quality of instructors, is of secondary importance.

Conclusions, Caveats, and Recommendations

The introductory business statistics course (STATS250) at Suffolk University was taught in 22 sections in 2013, of which eight followed a flipped hybrid model. Based on data from a common final exam, an end-of-the-semester questionnaire completed by students, and matched data from the Registrar's office, and using a variety of techniques, some of which sought to correct for selection bias, we were able to assess whether hybrid sections were associated with better outcomes.

The results are clear and consistent. Controlling for other observable factors, students in the hybrid courses performed better on the common final exam only in spring 2013; however, their semester grades, rating of the course, and newfound enthusiasm for statistics did not differ significantly from those of their peers in traditional classes in that semester.

The evidence of stronger outcomes for hybrid sections was weaker in fall 2013 (when half of the sections were taught in this manner) than in spring 2013: none of the outcome measures were significantly different between hybrid and traditional classes. Presently, our best assessment is that hybrid sections in introductory business statistics do no harm, at least as measured by a relatively standard set of outcomes, but they do not yield better outcomes than traditional courses either. These results are consistent with those of Terry (2007) who compares student performance in traditional, online, and hybrid formats of graduate Master of Business Administration (MBA) courses. He finds that the hybrid format maintains the high quality and student satisfaction associated with traditional F2F instruction. Some might argue that in the not-so-distant future the hybrid format may outperform the traditional format as faculty gain experience in this type of environment, and as further technological advances improve mode efficiency.

Cost Savings?

If the outcomes of hybrid sections are not demonstrably better than those of traditional sections, the case for using the hybrid model rests entirely on the potential for cost savings. We have assembled some indicative numbers in Table 7.

The assumptions are set out at the bottom panel of the table. The most important cost drivers are the number of students per section (28), the cost of adjunct faculty (total of $5,000 per course per semester), the cost of full-time faculty ($80,000 per year, plus benefits including sabbaticals, and a five-course teaching load per year), and the mix of adjunct and full-time faculty. The middle column approximates the situation at Suffolk University, where costs are probably in line with many mid-range universities in the United States, especially those that employ a mix of part- and full-time faculty.

We estimate the marginal cost of teaching introductory business statistics to one student for one semester to be $376 in traditional classes, $289 if half the courses are hybrids, and $211 if all the sections are taught in hybrid format. The savings from moving to hybrids calculated here (i.e. $164 per student per course) include lower salaries and benefits, but also reductions in the cost of providing classroom and office space. The totals here may look low – the average instructional cost per student of a course at Suffolk University is just over $1,000, and about $2,800 when support and overhead costs are factored in – but the modest unit costs reflect the relatively high average class size and substantial use of adjunct faculty in introductory business statistics classes. Note that we assume that the move from traditional to hybrid courses would be associated with a higher proportion of full-time faculty teachers, up from 25% to 33% in our illustrative example; this reflects the practice of asking full-time faculty first to teach hybrid sections.

Over time, the introduction of hybrids may also lead to savings in some of the costs of academic and institutional support – the human resources office could be smaller, fewer security personnel may be needed, heating and lighting costs would be lower, and so on. Based on publicly-available financial data, and assuming that half of “academic support” and “institutional support” costs are adjustable in this way, we estimate these costs to be 63% of “instructional costs” at Suffolk University, and this is reflected in our “marginal cost including overhead” panel in the middle of Table 7. If ten percent of undergraduate courses at Suffolk University were taught as hybrids rather than in the traditional manner, the estimated annual savings would be $1.2 million annually, or 0.6% of total operating expenditures – certainly a worthwhile saving, but not an overwhelming one. Ironically, the potential savings are substantially greater in institutions that rely almost exclusively on full-time faculty to teach introductory courses such as STATS250.

Based on our findings, it is reasonable to recommend that hybrid courses be used in teaching introductory business statistics: hybrids do no harm, are sustainable, may be attractive to some students, and reduce costs somewhat. Not every student thrives on the flipped hybrid model, so there is a strong case for continuing to offer a mix of traditional and hybrid courses.

References

Ary, E. J., & Brune, C. W. (2011). A comparison of student learning outcomes in traditional and online personal finance courses. *MERLOT Journal of Online Learning and Teaching 7*(4), 465-74.

Bowen, W. G., Chingos, M. M., Lack, K. A., & Nygren, T. I. (2012). Interactive Learning Online at Public Universities: Evidence from Randomized Trials. *Ithaka S+R*, May 22.

Carnegie Foundation. (2014). Summary Tables: Undergraduate Instructional Program Classification. http://classifications.carnegiefoundation.org/summary/ugrad_prog.php (accessed February 17, 2014).

CUNY Hybrid Initiative. 2014. Flipping the Classroom. http://hybrid.commons.gc.cuny.edu/teaching/additional-resources/flipping-the-classroom/ (accessed February 16, 2014)

Driscoll, A., Jicha, K., Hunt, A. N., Tichavsky, L., & Thompson, G. (2012). Can Online Courses Deliver In-Class Results? A Comparison of Student Performance and Satisfaction in an Online Versus a Face-to-Face Introductory Sociology Course. *Teaching Sociology 40*(4), 312-31.

Ettner, S. (Undated). *Methods for Addressing Selection Bias in Observational Studies*. Division of General Internal Medicine and Health Services Research, UCLA.

Figlio, D. N., Rush, M., & Yin, L. (2010). Is it Live or is it Internet? Experimental Estimates of the Effects of Online Instruction on Student Learning. *National Bureau of Economic Research Working Paper* No. 16089, http://www.nber.org/papers/w16089 (accessed January 28, 2014).

Haughton, D., & Haughton, J. (2011). *Living Standards Analytics: Development through the Lens of Household Survey Data*, Springer.

Jaggia, S., & Kelly, A. (2012). *Business Statistics: Communicating with Numbers*, McGraw-Hill/Irwin.

Lack, K A. (2013). Current Status of Research on Online Learning in Postsecondary Education. *Ithaka S+R*, March 21.

Lam, M, (2009). Effectiveness of web-based courses on technical learning. *Journal of Education for Business, 6*, 323-331.

Means, B., Toyama,Y., Murphy, R., Bakia, M., & Jones, K. (2010). Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies. U.S. Department of Education, Office of Planning, Evaluation, and Policy Development, Washington, DC.

NCES (National Center for Education Statistics). (2014). *Digest of Education Statistics* 2013. U.S. Department of Education, Institution of Education Sciences. http://nces.ed.gov/programs/digest/ (accessed February 22, 2014).

Rosenbaum, P., & Rubin, D. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, *70*(1), 41-55.

Sebastianelli, R., & Tamimi, N. (2011). Business Statistics and Management Science Online: Teaching Strategies and Assessment of Student Learning. *Journal of Education for Business*, *86*, 317-325.

Terry, N., (2007). Assessing Instruction Modes for Master of Business Administration (MBA) Courses. *Journal of Education for Business, 4,* 220-225.

U.S. News. (2014). College Compass. http://colleges.usnews.rankingsandreviews.com/best-colleges/suffolk-university-2218 and http://colleges.usnews.rankingsandreviews.com/best-colleges/rankings/regional-universities-north/data/spp+50 (accessed February 26, 2014).

Wilson, D., & Allen, D. (2014). Success rates of online versus traditional college students, *Research in Higher Education 14*, 1-8, http://www.aabri.com/manuscripts/11761.pdf (accessed March 31, 2014).

Table 1. Is There Response Bias?

	Spring 2013		Fall 2013	
	Number	Percentage	Number	Percentage
Listed on final exam roster	279	100.0	309	100.0
- Did not take final exam	-11*		-11*	
- Took final exam, but did not complete survey	-9*		-42*	
= Potential usable full responses	259	92.8	257	82.8
- Completed survey, but no id (or id unmatched)	-26*		-25	
= Observations with full responses	233**	83.5	232**	74.8

	Responders**	Non-responders*	Responders**	Non-responders*
Final exam score	69.2	59.4	66.9	65.7
Final grade	2.58	1.95	2.34	2.07
Admissions rating (bottom=1, top=10)	5.76	4.76	5.47	4.82
In an honors program (yes=1)	0.18	0.02	0.13	0.09
Age (in years)	20.67	20.88	20.52	20.93
Male? (yes=1)	0.49	0.65	0.42	0.45
Sophomore (yes=1)	0.42	0.48	0.66	0.52
Junior (yes=1)	0.23	0.22	0.22	0.31
Senior (yes=1)	0.07	0.07	0.09	0.16
In College of Arts and Sciences (yes=1)	0.26	0.20	0.22	0.23
Number of credits taken in semester (min 4, max 20)	15.30	14.07	15.55	15.51
Cumulative GPA (min 0, max 4)	3.17	2.87	3.11	3.04
GPA used at time of admissions	3.04	2.68	3.00	2.89
Major:				
Accounting	0.12	0.07	0.09	0.10
Global business	0.09	0.07	0.11	0.09
CAS science	0.12	0.13	0.15	0.17
Economics	0.04	0.02	0.02	0.01
Business	0.14	0.13	0.12	0.12
Entrepreneurship	0.10	0.11	0.06	0.14
Finance	0.11	0.11	0.14	0.08
Management	0.10	0.15	0.10	0.14
Marketing	0.07	0.11	0.16	0.06

Table 2. Checking Variation in Outcomes Over Time

	Fall 2012	Spring 2013	Fall 2013
Common final exam: Grades (%)			
Mean	**67.07**	**68.14**	**66.64**
Standard deviation	21.98	20.10	17.21
t-statistic vs. fall 2012	*n.a.*	*0.59*	*
Course grades (scale 0-4)			
Mean	**2.62**	**2.50**	**2.28**
Standard deviation	1.18	1.24	1.18
t-statistic vs. fall 2012	*n.a.*	*-1.20*	*-3.47*
Rating of course: low(1)-high(5)			
Mean	**2.99**	**2.97**	**2.77**
Standard deviation	1.19	1.11	1.13
t-statistic vs. fall 2012	*n.a.*	*-0.19*	*-2.12*
Raised interest: no(1)-a lot(5)			
Mean	**2.93**	**2.78**	**2.64**
Standard deviation	1.28	1.30	1.20
t-statistic vs. fall 2012	*n.a.*	*-1.22*	*-2.55*

Note: * Final exam in fall 2013 not strictly comparable with the one given in fall 2012 and spring 2013.

Table 3. Testing for the Impact of Hybrid Courses: Uncontrolled Differences

	Hybrid	Non-hybrid	Difference	p-value of t-test	p-value of χ^2 test*	No. of observations
Common final exam: Grades (%)						
Spring 2013	75.11	66.84	**8.26**	**0.01**		267
Fall 2013	68.49	65.19	**3.30**	**0.10**		298
Course grades (scale 0-4)						
Spring 2013	2.53	2.49	0.03		0.08	289
Fall 2013	2.24	2.31	-0.07		0.27	303
Rating of course: low(1)-high(5)						
Spring 2013	3.17	2.93	0.24		0.08	231
Fall 2013	2.81	2.73	0.07		0.98	250
Raised interest: no(1)-a lot(5)						
Spring 2013	3.26	2.68	**0.59**		**0.05**	231
Fall 2013	2.63	2.65	-0.02		0.25	253

Note: * The χ^2 test checks whether the distribution of cells in a cross-tabulation differs between hybrid and traditional classes.

Table 4. Regression Estimates: Performance on Common Final Exam

	Spring 2013		Fall 2013		Mean values	
	Coefficient	**p-value**	**Coefficient**	**p-value**	**Spring 2013**	**Fall 2013**
Dependent variable: grade on common final exam					67.9	66.6
Hybrid section (yes=1)	**9.511**	**0.00**	3.279	0.16	0.15	0.42
Prior statistics (none=1, college=5)	1.729	0.13	**3.159**	**0.06**	1.67	1.66
Course expected to be hard (v. easy=1, very difficult=5)	**-2.715**	**0.09**	-0.680	0.45	3.74	3.72
Wanted or needed to learn subject (not imp=1, v.imp=4)	3.193	0.18	0.917	0.70	0.40	0.37
Max higher education of parent (years)	0.301	0.63	-0.187	0.75	14.95	14.59
Expected higher education (Assoc.=1, Post-Masters=5)	3.299	0.11	-0.520	0.77	3.82	3.67
Number of credits taken in semester (min 4, max 20)	0.714	0.54	0.759	0.37	15.07	15.53
Hours per week of coursework (outside class)	**-0.335**	**0.02**	0.053	0.75	10.81	10.76
Hours per week worked for pay	**-0.169**	**0.07**	-0.090	0.43	12.18	12.17
Live with parents/family? (yes=1)	3.543	0.25	-2.815	0.41	0.30	0.34
Live independently? (yes=1)	**6.928**	**0.08**	-2.324	0.43	0.40	0.45
English is main language (yes=1)	-3.391	0.27	-3.897	0.19	0.56	0.57
Chinese speaker (yes=1)	**11.929**	**0.01**	1.376	0.88	0.06	0.07
Spanish speaker (yes=1)	-6.588	0.23	0.379	0.94	0.10	0.10
Used SPSS or Excel (yes=1)	-2.332	0.26	9.581	0.11	0.42	0.72
Technical computer problems? (Never=1, very often=4)	-1.971	0.56	-1.170	0.72	0.14	0.09
Admissions rating (bottom=1, top=10)	1.301	0.28	1.301	0.19	5.57	5.31
GPA used at time of admissions	1.023	0.73	1.857	0.51	2.95	2.98
Cumulative GPA (min 0, max 4)	4.762	0.26	**7.018**	**0.04**	3.08	3.09
In an honors program (yes=1)	5.848	0.22	**6.158**	**0.08**	0.14	0.11
Age (in years)	-0.357	0.54	-0.030	0.97	20.69	20.61
Male? (yes=1)	0.512	0.82	0.073	0.97	0.48	0.40
Sophomore (yes=1)	1.214	0.70	**-17.103**	**0.05**	0.39	0.58
Junior (yes=1)	-1.329	0.76	**-18.341**	**0.07**	0.21	0.22
Senior (yes=1)	-1.153	0.83	-17.519	0.21	0.06	0.10
Major:*						
Accounting	**10.254**	**0.02**	4.066	0.44	0.10	0.09
Global business	3.670	0.59	-2.662	0.55	0.09	0.10
CAS science	**10.983**	**0.04**	11.383	0.17	0.11	0.14
Economics	5.237	0.39	11.947	0.26	0.03	0.02
Entrepreneurship	3.424	0.59	0.455	0.92	0.09	0.08
Finance	-0.728	0.91	3.275	0.47	0.10	0.11
Management	6.509	0.38	2.814	0.67	0.10	0.10
Marketing	-1.638	0.75	1.322	0.76	0.07	0.12
Other major	7.921	0.12	**14.080**	**0.00**	0.11	0.06
Intercept	25.805	0.36	28.688	0.26		
Memo items:						
Number of observations	214		208			
R squared	0.40		0.35			

Note: * Omitted major is "Business".

Table 5. Propensity Score Equation Estimates (Probit)

	Spring 2013		Fall 2013	
	Coefficient	p-value	Coefficient	p-value
Dependent variable: hybrid section (yes=1)				
Prior statistics (none=1, college=5)	-0.002	0.99	**-0.257**	**0.03**
Course expected to be hard (very easy=1, very difficult=5)	-0.074	0.55	0.166	0.13
Wanted or needed to learn subject (not imp=1, v.imp=4)	0.372	0.14	0.073	0.71
Max higher education of parent (years)	0.018	0.73	0.025	0.56
Expected higher education (Assoc.=1, Post-Masters=5)	0.021	0.91	-0.017	0.88
Number of credits taken in semester (min 4, max 20)	-0.061	0.39	0.107	0.13
Hours per week of coursework (outside class)	0.014	0.47	-0.004	0.85
Hours per week worked for pay	0.002	0.85	0.000	0.96
Live with parents/family? (yes=1)	**0.987**	**0.01**	**0.564**	**0.06**
Live independently? (yes=1)	**0.730**	**0.06**	0.008	0.98
English is main language (yes=1)	-0.255	0.42	0.211	0.42
Chinese speaker (yes=1)	-0.592	0.26	**-1.780**	**0.00**
Spanish speaker (yes=1)	**-1.174**	**0.02**	0.317	0.37
Admissions rating (bottom=1, top=10)	0.052	0.61	-0.015	0.85
GPA used at time of admissions	-0.234	0.39	0.172	0.48
Cumulative GPA (min 0, max 4)	-0.091	0.64	-0.237	0.24
In an honors program (yes=1)	-0.368	0.40	0.508	0.12
Age (in years)	0.003	0.97	0.008	0.88
Male? (yes=1)	0.358	0.22	0.048	0.83
Sophomore (yes=1)	0.239	0.50	-0.737	0.45
Junior (yes=1)	-0.109	0.83	-0.889	0.37
Senior (yes=1)	-0.916	0.22	-0.849	0.41
Major:*				
Accounting	**1.143**	**0.04**	**-0.949**	**0.04**
Global business	**1.212**	**0.04**	-0.511	0.23
CAS science	0.880	0.17	-0.595	0.21
Economics	0.346	0.66	-1.047	0.19
Entrepreneurship	**1.072**	**0.07**	-0.776	0.12
Finance	0.730	0.21	-0.651	0.11
Management	0.346	0.61	-0.336	0.44
Marketing	0.849	0.19	-0.540	0.18
Other major	0.614	0.32	-0.944	0.12
Intercept	-1.121	0.63	-1.166	0.58
Memo items:				
Number of observations	215		209	
Pseudo R squared	0.20		0.15	

Note: Omitted major is “Business”.

Table 6. Alternative Measures of Impact of Hybrid Courses in Statistics

	Spring 2013		Fall 2013		Notes: k / n / R^2	
	Difference	p-value	Difference	p-value	Spring 2013	Fall 2013
Common final exam: Grades (%)						
Uncontrolled difference	**8.26**	**0.014**	3.30	0.100		
Regression, large model*	**9.51**	**0.001**	3.28	0.164	35 / 214 / .41	35 / 208/ .35
Regression, stepwise exclusion (if p>0.2)*	**10.10**	**0.001**	3.02	0.126	17 / 214/ .39	13 / 208/ .33
Regression, stepwise inclusion (if p<0.2)*	**8.21**	**0.024**	3.26	0.101	15 /214/ .34	10 / 208/ .31
Propensity Score Matching, ATT, nrst neighbor	**9.05**	**0.043**	5.44	0.954		
Direct matching, ATT	**12.36**	**0.005**	**5.29**	**0.053**		
Treatment regression	**25.60**	**0.021**	-0.77	0.950	p(λ) = 0.153	p(λ) = 0.751
Course grades: scale of low (0) to high (4)						
Uncontrolled difference	0.031	0.881	-0.068	0.617		
Regression, large model*	0.098	0.503	0.084	0.578	35 / 215/ .41	35 / 209/ .46
Regression, stepwise exclusion (if p>0.2)*	not included		not included		15 / 215/ .35	13 / 209/ .43
Regression, stepwise inclusion (if p<0.2)*	not included		not included		11 / 215/ .33	11 / 209/ .40
Propensity Score Matching, ATT, nrst neighbor	0.114	0.727	0.201	0.792		
Direct matching, ATT	0.15	0.155	0.235	0.188		
Treatment regression	0.874	0.189	0.124	0.867	p(λ) = 0.254	p(λ) = 0.942
Rating of course: “much worse” (1) to “much better” (5)						
Uncontrolled difference	0.241	0.204	0.070	0.629		
Regression, large model*	0.094	0.280	0.067	0.639	35 / 213/ .34	35 / 206/ .36
Regression, stepwise exclusion (if p>0.2)*	not included		not included		21 / 213/ .32	12 / 206/ .32
Regression, stepwise inclusion (if p<0.2)*	not included		not included		16 / 213/ .32	14 / 206/ .30
Propensity Score Matching, ATT, nrst neighbor	0.099	0.617	0.099	0.664		
Direct matching, ATT	0.421	0.136	0.225	0.279		
Treatment regression	0.73	0.275	0.490	0.529	p(λ) = 0.385	p(λ) = 0.589
Raised interest in subject: “not at all” (1) to “a great deal” (5)						
Uncontrolled difference	**0.585**	**0.008**	-0.019	0.903		
Regression, large model*	**0.329**	**0.038**	-0.124	0.432	35 / 213/ .29	35 / 209/ .29
Regression, stepwise exclusion (if p>0.2)*	**0.334**	**0.041**	not included		18 / 213/ .26	13 / 209/ .25
Regression, stepwise inclusion (if p<0.2)*	**0.424**	**0.005**	not included		13 / 213/ .25	13 / 209/ .25
Propensity Score Matching, ATT, nrst neighbor	-0.044	0.393	-0.274	0.119		
Direct matching, ATT	**0.711**	**0.021**	0.063	0.765		
Treatment regression	**1.681**	**0.041**	0.873	0.360	p(λ) = 0.111	p(λ) = 0.294

Note: Difference measures score for hybrid sections minus score for non-hybrid sections. “Not included” means that the variable indicating whether a student was in a hybrid course was not sufficiently statistically significant enough to be included in the final version of the equation. * cluster robust estimation. k / n / R^2 refers to number of included variables, number of observations, and R-squared respectively. p(λ) gives the p-value for a test of whether a 2-equation treatment regression model is preferable to a single-equation model.

Table 7. Illustrative Cost Savings from Using Hybrid Courses in Statistics

	All adjuncts	Mix of faculty	All f.t. faculty	
Marginal cost, excluding most overhead				
Cost/student/semester, all traditional classes, $	228	**376**	819	
Cost/student/semester, half hybrid, half traditional, $	160	**289**	603	
Cost/student/semester, all hybrid classes, $	114	211	409	
Memo: % gain going to half hybrid	30	**23**	26	
Memo: Cost savings/student/semester for hybrid, $	114	**164**	409	
Marginal cost, including overhead				
Cost/student/semester, all traditional classes, $	340	580	1,303	Add 63% to "instructional support" to cover half of "academic support" and half of "institutional support"
Cost/student/semester, half hybrid, half traditional, $	244	454	966	
Cost/student/semester, all hybrid classes, $	170	329	651	
Memo: Cost savings/student/semester for hybrid, $	170	252	651	
Overall effect on budget:				
5,800 u/g; 10% courses hybrid; $m	0.84	1.24	3.21	
Memo: as % of total operating expenses		0.59		
Assumptions				
Students per section	28	28	28	
Cost per section taught, adjunct faculty, $	5,000	5,000	5,000	$5,000 per section, including benefits
Cost per traditional section taught, f.t. faculty, $	21,538	21,538	21,538	$80,000 p.a. plus benefits, 5 classes per year
Mix: proportion taught by f.t. faculty				
Non-hybrid sections	0.00	0.25	1.00	
Hybrid sections	0.00	0.33	1.00	
Cost per room per class period per semester, $	180	180	180	$1,200/month/classroom, 25 periods/week
Office space per faculty member per section, $	1,200	1,200	1,200	$500/month, 5 classes per faculty member p.a.

ENDNOTES

[1] Students were first invited to participate in the Bowen et al. study, which provided monetary or other incentives; if they agreed, they were assigned randomly to F2F or hybrid sections of the course. Thus the randomization was conditional on students agreeing to participate. Of the 3,046 students enrolled in the course at the six campuses, 605 participated in the study, and of these, 292 were assigned to traditional sections and the remaining 313 to hybrid classes. The selection of instructors to teach the hybrid sections was not random either.

[2] The catalog description of STATS 250 is as follows:

> Application of statistical analysis to real-world business and economic problems. Topics include data presentation, descriptive statistics including measures of location and dispersion, introduction to probability, discrete and continuous random variables, probability distributions including binomial and normal distributions, sampling and sampling distributions, statistical inference including estimation and hypothesis testing, simple and multiple regression analysis. The use of computers is emphasized throughout the course. Normally offered each semester.

Before enrolling in the course, students must have successfully completed a course in college-level mathematics.

[3] Using the Carnegie classification, Regional Universities offer a full range of undergraduate programs and some master's programs, but few doctoral programs.