

Online Learning in Postsecondary Education: A Review of the Empirical Literature (2013–2014)

March 11, 2015

D. Derek Wu





Ithaka S+R is a strategic consulting and research service provided by ITHAKA, a not-for-profit organization dedicated to helping the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways. Ithaka S+R focuses on the transformation of scholarship and teaching in an online environment, with the goal of identifying the critical issues facing our community and acting as a catalyst for change. JSTOR, a research and learning platform, and Portico, a digital preservation service, are also part of ITHAKA.

Copyright 2015 ITHAKA. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of the license, please see <http://creativecommons.org/licenses/by-nc/4.0/>.

ITHAKA is interested in disseminating this brief as widely as possible. Please contact us with any questions about using the report: research@ithaka.org.

Contents

Introduction	3
Prior Literature and Context.....	4
Criteria for Studies	6
Findings.....	7
Threats to Validity.....	13
Avenues for Further Research	15
Conclusion.....	18
Description of Studies	19
Bibliography of Studies	43

Introduction

Courses that incorporate online learning are increasingly a fact of life for American college and university students.¹ The share of postsecondary students in the United States who took at least one online course has increased every year for the past decade—to a high of 34% in the fall semester of 2012.² Even as the prevalence of online learning continues to grow, however, there remains a dearth of rigorous research done on the learning outcomes associated with online and hybrid learning.³

Building on a 2013 Ithaka S+R report,⁴ this paper examines in depth the status of the research on learning outcomes associated with online and hybrid courses since 2013.⁵ Among the twelve studies evaluated for this report, only one employs a randomized controlled trial, and two more utilize quasi-experimental research designs to estimate the causal impact of online delivery formats on student outcomes. These more methodologically robust studies find that students taking online or hybrid courses generally performed no differently (or only marginally worse) than their peers in face-to-face sections, a finding consistent with prior studies of similar rigor. Yet, for the vast majority of studies analyzed, there has been little progress in addressing the methodological shortcomings underscored by Lack (2013) and others. Furthermore, there remains an insufficient body of evidence regarding the cost-effectiveness of online and technology-enhanced learning, one of the key research needs identified by Lack (2013).

This report begins by discussing prior efforts to survey the literature on online learning. Next, it offers the first summary and analysis of the research produced since 2013. The main body of the report concludes by examining potential avenues for improvement, emphasis, and future research. Finally, a “Description of Studies” section contains detailed, systematic summaries of each of the post-2013 studies assessed.

¹ I thank Martin Kurzweil, Johanna Brownell, and Christine Mulhern for helpful comments and discussions.

² Allen, Elaine I. and Jeff Seaman. “Grade Change: Tracking Online Education in the United States.” Babson Survey Research Group and Quahog Research Group (2014). <http://www.onlinelearningsurvey.com/reports/gradechange.pdf>.

³ “Hybrid” courses refer to those that contain both an online and a face-to-face component, although there exists substantial variation in how these courses are specifically structured. They can also be referred to interchangeably as “blended” courses.

⁴ Lack, Kelly A. “Current Status of Research on Online Learning in Postsecondary Education.” Ithaka S+R (2013). <http://sr.ithaka.org/sites/default/files/reports/ithaka-sr-online-learning-postsecondary-education-may2012.pdf>.

⁵ In particular, this literature review includes only studies that were released or published during and after January 2013 in order to align it with the latest iteration of Lack’s literature review, which covered studies that were published no later than December 2012.

Prior Literature and Context

There exist a number of studies published prior to 2013 that have examined the impact of online and hybrid courses on learning outcomes, but few have been methodologically rigorous enough to provide conclusive evidence about the effectiveness of online or hybrid learning. A widely-cited 2010 meta-analysis prepared by Means et al. for the U.S. Department of Education combs findings from 1,132 studies on online learning published between 1996 and 2008 and finds that very few of these studies attempted to assess the causal impact of online delivery formats on learning outcomes.⁶ In fact, the authors observe that only 45 studies (less than 4% of the total) directly compare web-based instruction to face-to-face instruction, employ experimental or quasi-experimental research designs, and focus on objective measures of student learning. The authors' meta-analysis of those few studies indicates that students taking fully online courses performed marginally better than their counterparts in face-to-face sections, whereas students who took courses in a hybrid format performed significantly better than those in face-to-face sections.

Another extensive literature review carried out by Lack (2013) surveys a different set of studies and finds little evidence across multiple outcome measures that online or hybrid learning is more or less effective than face-to-face learning. However, one of the studies surveyed by Lack (2013) proves to be a notable exception to these overall trends, finding via an instrumental variable approach that online courses taken by community college students were associated with significantly negative estimates for course persistence and grade (Xu and Jaggars (2011)).⁷ However, the fact that this particular study focuses on a different population of students than other studies examining four-year institutions makes it difficult to compare them directly.

Two of the more methodologically robust studies released prior to 2013 that were not evaluated in either of these literature reviews were those of Figlio et al. (2010) and

⁶ Means, Barbara, Yukie Toyama, Robert Murphy, Marianne Bakia, and Karla Jones. "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 (2010). <http://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf>.

⁷ Xu, Di and Shanna Smith Jaggars. "Does Course Delivery Format Matter? Evaluating the Effects of Online Learning in a State Community College System Using Instrumental Variable Approach." Community College Research Center Working Paper No. 31, Teachers College, Columbia University (2011). https://aefpweb.org/sites/default/files/webform/Online_learning_using_instrumental_variable_approach.pdf. A related version of this paper was subsequently published in the *Economics of Education Review*, Vol. 37 (2013).

Bowen et al. (2012).⁸ Figlio et al. (2010) randomly assign students in an introductory microeconomics course to lectures offered live or in an online setting. They find statistically insignificant evidence that the live format produced more favorable impacts than the web-based format, although these effects were significant for students who were lower-achieving, male, and Hispanic.⁹ Bowen et al. (2012) also undertake a large-scale study between Fall 2010 and Spring 2012 that randomized more than 600 students on six public college campuses into hybrid (with some face-to-face instruction) or purely face-to-face versions of an introductory statistics course and find no significant differences in learning outcomes between students in the two course types.¹⁰

In sum, the prior literature generally indicates that online and hybrid course formats produce outcomes that are not significantly different from those in face-to-face formats. However, given how few studies employ methodologies that yield a causal inference (or, for that matter, employ sufficient controls), much more rigorous research is needed to ensure that these results are robust to various specifications and settings. The prior literature also lacks sufficient evidence regarding how the effects of online and hybrid courses vary across different student subgroups and extend over longer periods of time, as well as careful analyses of the costs associated with these delivery formats. While a few studies included in this review employ greater methodological rigor and comprehensiveness, the majority of studies still fall short in their efforts to fill in the gaps left by the prior literature—particularly those related to the cost implications of online and hybrid delivery formats.

⁸ Lack (2013) mentions both studies but does not include them in her formal review.

⁹ Figlio, David, Mark Rush, and Lu Yin. "Is it Live or Is It Internet? Experimental Estimates of the Effects of Online Instruction on Student Learning." NBER Working Paper 16089 (2010). <http://www.nber.org/papers/w16089>. A version of this report has since been published in the *Journal of Labor Economics*, Vol. 31, No. 4 (2013).

¹⁰ Bowen, William G., Matthew M. Chingos, Kelly A. Lack, and Thomas I. Nygren. "Interactive Learning Online at Public Universities: Evidence from Randomized Trials." Ithaka S+R (2012). <http://sr.ithaka.org/sites/default/files/reports/sr-ithaka-interactive-learning-online-at-public-universities.pdf>. A version of this report has since been published in the *Journal of Policy Analysis and Management*, Vol. 33, Issue 1 (Winter 2014).

Criteria for Studies

For the sake of consistency, this literature review relies on the same criteria as those introduced by Lack (2013) to select studies for inclusion. To be considered, studies had to:

1. Compare at least one face-to-face section to at least one hybrid or fully online section.
2. Examine objective learning outcomes or measures of academic performance that are not self-reported.
3. Involve at least one undergraduate, for-credit college course offered outside of a continuing education program.
4. Take place in the United States or in a country with a comparable culture and higher education system.
5. Be authored by someone who is not a current student.

Twelve studies were identified that met these criteria, with three published in 2013 and nine in 2014. All focus on public institutions, all but one focus on students at four-year institutions, and only one study examines students outside the United States. Two studies examine courses in multiple subjects, and the rest focus on courses in a single subject (with four in economics, two in statistics, one in information systems, one in interdisciplinary writing, one in psychology, and one in stress management). The sample sizes range from 75 to 40,000 students, although the median study comprises several hundred observations. The “Description of Studies” section at the end of this report includes a detailed summary of and commentary on each of the twelve studies.

Findings

The studies included in this review are categorized by research design: experimental and quasi-experimental studies whose methodologies permit inferences of causality, descriptive studies with robust controls, and descriptive studies without robust controls.¹¹ The three studies that employ randomization or quasi-experimental strategies find that students taking online or hybrid courses performed slightly worse to no differently than their peers taking traditional face-to-face courses, though there is some variance across specifications and subgroups. The descriptive studies that incorporate control variables find that online and hybrid courses were generally associated with lower learning outcomes. Finally, a majority of the six studies that employ strictly observational analyses indicate that students in online and hybrid formats performed no worse and, in some cases, better than their counterparts in face-to-face sections.

Experimental and Quasi-Experimental Studies

Joyce et al. (2014) is the only included study to employ a randomized controlled trial, assigning 725 students randomly to one of four sections—two “compressed” and two traditional—of an introductory microeconomics course. Across all performance measures, the authors find that students in the traditional format performed better than those in the hybrid format, although performance converged as the semester progressed. Furthermore, high-performing students performed equally well regardless of delivery method, whereas students with average and lower levels of expected performance performed worse in the hybrid sections.

Kwak et al. (2015)¹² utilize a quasi-experimental research design, replacing two weeks of lectures in the middle of a semester-long, face-to-face introductory statistics course with a blended format that cut class time in half and made additional online materials

¹¹ Ithaka S+R also undertook a large-scale study in 2013 that analyzed student outcomes associated with hybrid formats in multiple universities across the University System of Maryland. This study conducted seven side-by-side comparisons of student outcomes in hybrid sections with outcomes in traditionally taught courses, controlling for a myriad of student background characteristics. It finds that students in the hybrid formats performed as well or even slightly better than their face-to-face peers on course pass rates and learning assessments, a result that held across subjects and student subgroups. Furthermore, the statistical insignificance of this result is consistent with the findings of the experimental and quasi-experimental studies included in this literature review. However, to maintain objectivity, this literature review chooses not to include this particular study in its formal review. For the full paper and its results, see Griffiths, Rebecca, Matthew Chingos, Christine Mulhern, and Richard Spies. “Interactive Online Learning on Campus: Testing MOOCs and Other Platforms in Hybrid Formats in the University System of Maryland.” Ithaka S+R (2014). http://sr.ithaka.org/sites/default/files/reports/S-R_Interactive_Online_Learning_Campus_20140716.pdf.

¹² Although this article was not formally published until March 2015, it was first published online in December 2014—within the timeframe that this literature review covers.

available for out-of-class use. Using a difference-in-differences identification strategy, the authors find that the blended format had no effect on students' quiz scores, although the effects were negative for male students and positive for female students. The other quasi-experimental study included in this review, Olitsky and Cosgrove (2014), employs a propensity score matching technique to control for bias stemming from non-random selection into a hybrid delivery format. The authors find that the hybrid section was associated with lower student results in several outcome measures, although these effects are statistically insignificant for nearly all measures.

Descriptive Studies with Robust Controls

Three studies employ multivariate regression analyses that control for student- and instruction-related factors correlated with selection of the delivery format and eventual learning outcomes. Two of these studies conduct descriptive analyses in settings with some characteristics of a controlled trial. Burns et al. (2013) compare student outcomes associated with being in a face-to-face, online, or hybrid version of an information systems course and find that face-to-face students received higher course grades than those in the other delivery formats. However, students in the online and hybrid sections of the information systems course outperformed their face-to-face peers in a subsequent business course. Verhoeven and Rudchenko (2013) examine an undergraduate introductory microeconomics course in which 51 students participated in a hybrid section and 24 students participated in a face-to-face section, and find that the hybrid format was associated with lower test scores than the face-to-face format. In a third study, Xu and Jaggars (2014) use an administrative dataset covering 40,000 community college students in Washington State over a five-year period. Controlling for various fixed effects and student- and instruction-related covariates, they find that students, especially those with lower achievement levels, received lower grades in online courses and were less likely to persist in online sections.

Descriptive Studies without Robust Controls

The remaining six studies in this review employ descriptive analyses that did not incorporate control variables, instead relying on purely observational comparisons between delivery formats. Nonetheless, these studies attempt to standardize their treatment (online/hybrid) and control (face-to-face) groups using the same instructors, course materials, and/or class structures. Jones and Long (2013) compare final course grades over ten semesters of a business mathematics course and, in the majority of semesters analyzed, find no significant differences between mean scores for face-to-face and online students. Carmichael et al. (2014) examine learning outcomes associated with online, hybrid, and face-to-face versions of an interdisciplinary writing course and

observe that improvements in writing skills between the beginning and end of the semester were largest for students in the hybrid and online sections.

Fish and Kang (2014) compare outcome data from 119 students divided between online and face-to-face sections of a stress management course and find no significant differences between delivery formats in the students' average score on three exams given during the term. However, the authors find that students who took the course in the online format earned higher scores on the final exam and that the difference is statistically significant. Jorczak and Dupuis (2014) evaluate the outcomes of 104 students who took face-to-face and online sections of an introductory psychology course, finding that online students scored 25 percentile points higher on course exams than their face-to-face counterparts. Metzgar (2014) analyzes learning outcomes associated with hybrid and face-to-face formats of a managerial economics course for business majors and finds that students in the hybrid sections performed worse than their in-class peers on both in-class questions and exams. Finally, Tanyel and Griffin (2014) analyze student outcomes associated with 81 courses offered over a ten-year period that each included at least one online and face-to-face section. The authors find that students in the face-to-face sections earned higher grades than those in the online sections, although these differences only appear in the last three years of their analysis.

Study Snapshots

Table 1 summarizes the key characteristics and findings of these studies. While online and hybrid delivery formats have not been shown to be consistently more or less effective than face-to-face instruction across all studies, the results of the studies reviewed indicate that the rigor of the research framework utilized significantly impacts the kinds of results observed.

Table 1. Snapshot of Study Characteristics and Results

Study	Sample	Institution	Course*	Research Design	Outcome Measures	Results
Joyce et al. (2014)	725 total students randomized into two FTF and online sections	Baruch College, City University of New York	Principles of Microeconomics (FTF/H)	Randomized controlled trial	Scores on the midterm and final exams, final course grades	FTF students scored 3.2 percentage points higher than hybrid students on the midterm, but the differences on the final were half as large and no longer statistically significant; larger differences among students with lower performance levels
Kwak et al. (2014)	First-year business and economics students	University of Queensland (Australia)	Introductory statistics course (FTF/H)	Quasi-experimental	Scores on weekly quizzes	Hybrid format reduced quiz scores by 1.9 percentage points (after incorporating covariates) in OLS specification; in difference-in-differences method, hybrid learning caused insignificant increase in quiz score of 0.16 percentage points
Olitsky and Cosgrove (2014)	236 FTF students, 82 hybrid students	N/A	Principles of Microeconomics and Macroeconomics (FTF/H)	Quasi-experimental	Exam scores, grades on homework assignments and short-answer questions	Hybrid format associated with statistically insignificant decrease in exam score of 5.5 percentage points under propensity score matching (versus 0.065 points under OLS)
Burns et al. (2013)	109 FTF* students, 144 online students, 129 hybrid students	Midwestern land grant university	Computers and Information Systems (FTF/O/H)	Observational with controls	Final grades in course with intervention; grades in subsequent more advanced course	Face-to-face format associated with positive, significant effects on current course grade (via ordered probit models with control variables); Online and hybrid formats associated with positive, significant effects on subsequent course grade

Study	Sample	Institution	Course*	Research Design	Outcome Measures	Results
Verhoeven and Rudchenko (2013)	24 FTF students, 51 hybrid students	Large public university	Principles of Microeconomics (FTF/H)	Observational with controls	Composite test scores, completion rates	71% of hybrid students completed (versus 79% of FTF students); hybrid students scored 4.8 percentage points lower than FTF peers on exam scores
Xu and Jaggars (2014)	40,000 total students	Washington State community college system	Encompassed 500,000 unique course enrollments (FTF/O)	Descriptive with fixed-effects regression analysis	Final course grades, course persistence	Average effects of online format on course persistence and grade negative; stronger effects for men, black students, younger, and at-risk students
Carmichael et al. (2014)	Senior undergraduates	University of North Dakota	Writing Across Disciplines (FTF/O/H)	Observational	Rubric-based scores for writing, critical thinking, and integrative learning studies	Average scores increased in all 3 rubric areas in all sections; improvement statistically significant at 95% confidence level only for one online section and hybrid section
Fish and Kang (2014)	63 FTF students, 56 online students	Large, public university on the U.S. West Coast	Upper-division stress management course (FTF/O)	Observational	Scores on three exams	No significant differences between formats when averaging scores across all exams; however, students in online course scored 3.4 percentage points higher on third test (significant at 95% confidence level)
Jones and Long (2013)	267 FTF students, 178 online students (mostly freshmen)	Small, open-enrollment rural Appalachian college	Quantitative Business Analysis (FTF/O)	Observational	Final course grades	Across all semesters and students, average FTF course grades were 5 percentage points higher than online grades; difference disappeared after dropping first three semesters

Study	Sample	Institution	Course*	Research Design	Outcome Measures	Results
Jorczak and Dupuis (2014)	35 FTF students, 69 online students	Medium-size Midwestern public university	Introductory psychology course (FTF/O)	Observational	Scores on two multiple-choice exams	Online students scored an average of 74% on exams, compared to 67% for F2F students; difference statistically significant
Metzgar (2014)	80 students in each of 3 hybrid sections	Large Southern public university	Managerial Economics (junior-level) (FTF/H)	Observational	Accuracy on in-class clicker questions and exam scores	Hybrid students scored 60-70% on exams (versus 70-80% for FTF students) and 30-60% on clicker questions (versus 70-80% for FTF)
Tanyel and Griffin (2014)	3,355 FTF students in 132 sections and 2,266 online students in 94 sections	Southeastern regional university	81 different courses over 10 years (with 66 in the College of Arts and Sciences) (FTF/O)	Observational	Final course grades, withdrawal rates	30% of online students failed or withdrew from class (compared to 18% of F2F students); average GPAs earned by F2F students 0.15 higher than online peers (although difference driven by last 5 semesters)

* FTF abbreviation designates "Face-to-Face"

** In addition to course titles, this column contains information on the types of delivery formats analyzed with the following categorizations:

FTF/O (Face-to-Face vs. Online), FTF/H (Face-to-Face vs. Hybrid), FTF/O/H (Face-to-Face vs. Online vs. Hybrid)

Threats to Validity

Many of the included studies are vulnerable to methodological limitations that endanger the robustness of their results. This section describes a few of these threats to validity and their implications for the results of the studies reviewed.

Only one study randomly assigns students to face-to-face or online/hybrid delivery formats, and only two other studies employ a quasi-experimental identification strategy to address sample selection bias. The evidence supporting causal inferences about the impact of online and hybrid delivery on learning outcomes is therefore quite thin. Of the nine remaining studies, six fail to account for observed and unobserved differences in student and classroom settings between delivery formats—even though some of these studies (Fish and Kang (2014), Jones and Long (2014)) note that students in the online sections tended to be older and lower-achieving than their peers in the face-to-face sections. Omitting these factors threatens the internal validity of the “effects” of online and hybrid learning that these studies observe, as these characteristics are very likely correlated with both selection of the delivery format and eventual learning outcomes. Even in the three descriptive studies that do control for pre-existing differences in student characteristics, the authors do not account for other factors that may be impacted by selection of the delivery format and are correlated with course performance. Examples of such potentially collinear factors include the difficulty of and academic performance in concurrent courses, whether students attend school on a part-time or full-time basis, and employment status.¹³

Furthermore, many of the descriptive studies feature relatively small sample sizes, with three assessing fewer than 120 students—who are in turn subdivided into multiple delivery formats. Such small sample sizes result in larger standard errors and fewer statistically significant results. Additionally, none of the studies account for attrition bias, an oversight that very seriously threatens a study’s validity when course performance at the end of the semester serves as the dependent variable of interest. If attrition occurred selectively between sections—for example, if students with lower achievement levels were more likely to withdraw from online courses than face-to-face courses—then the end-of-semester indicators of learning outcomes are poor estimates of the true impact of a delivery format on student outcomes. Controlling for pre-existing student characteristics would not account for differing propensities to withdraw based on the delivery format of the course. Indeed, of the three studies that treat course

¹³ While one may argue that these factors may be “predicted” by pre-existing covariates, it is important to note that their association with delivery format is one in which they do not affect—but rather are affected by—the delivery format.

retention as a separate outcome, two find that withdrawal rates from online and hybrid courses were higher than the rates from face-to-face courses.

In addition, several descriptive studies that attempt to standardize certain features of courses in both delivery formats nevertheless allow differences across sections that threaten the generalizability of their results. For example, in two studies, students in one section but not the other were graded for attendance or participation. To the extent that those grades motivated students to attend or participate in class more than their peers in other sections—which in turn may have led to differential learning outcomes—the unaccounted-for heterogeneity in grading policies may have undermined efforts to isolate the impact of the delivery format. Moreover, in two of the studies reviewed, exams for the face-to-face section were given in an in-class, proctored setting, whereas exams for the online and hybrid sections were offered in a remote, un-proctored setting. While it makes practical sense to test students in a form consistent with the delivery format of their section, such differences in testing format call into question the comparability of results.

Finally, several studies do not adequately define or differentiate the types of online and hybrid courses they study, with two studies grouping fully online and hybrid sections under a single “online” umbrella. This “lumping” likely muddies their results, given that other studies have shown that online and hybrid delivery formats produce different learning effects in terms of magnitude, direction, and statistical significance. The imprecision in categorizing delivery formats also makes it more difficult to compare results across studies. For example, hybrid formats in one study may have featured extensive interaction between students in lieu of in-class lecturing (see, for example, Metzgar (2014)), whereas those in another study may have served merely as compressed, lecture-heavy versions of the face-to-face formats (see, for example, Joyce (2014)).

Avenues for Further Research

As the analysis of threats to validity makes clear, there remains a need for greater methodological rigor in the research on learning outcomes associated with online and hybrid instruction. At the same time, there are several related research questions that deserve more attention than they have received.

First, there exists a need for more rigorous research on the cost implications of online and hybrid instruction. None of the studies included in this literature review examine the effect of delivery formats on course costs, and yet several suggest that the potential cost reductions—or increases—associated with online and hybrid courses may be what ultimately drive the extent to which their results are actionable (Joyce et al. (2014), Kwak et al. (2015), Olitsky and Cosgrove (2014)). To support action on the ground, the research must address not only the effects of online and hybrid instruction on learning outcomes, but also their efficiency. Griffiths et al. (2014) and others observe that online and hybrid courses have higher fixed (start-up) costs than face-to-face delivery formats.¹⁴ As Cowen and Tabarrok (2014) and Bowen et al. (2012) have argued, however, the marginal costs associated with online and hybrid courses may be significantly less than those associated with face-to-face formats and should diminish every time that the online or hybrid section is offered.¹⁵ The challenge to studying the long-term cost effects in the field is that many efforts at online and hybrid instruction are curtailed before the high start-up costs can be amortized.¹⁶ A commitment to a sustained experiment, likely at some scale, is needed to test the productivity question.¹⁷

¹⁴ Griffiths et al. (2014), p. 47.

¹⁵ For further discussion on this, see Cowen, Tyler and Alex Tabarrok. “The Industrial Organization of Online Education.” *American Economic Review: Papers and Proceedings* 104, no. 5 (2014): 519-522. For illustrative simulations of the potential longer-term cost savings associated with a technology-enhanced course, see Appendix B of Bowen et al. (2012).

¹⁶ See Marcum, Deanna, Christine Mulhern, and Clara Samayoa. “Technology-Enhanced Education at Public Flagship Universities: Opportunities and Challenges.” *Ithaca S+R* (2014).
http://sr.ithaka.org/sites/default/files/reports/SR_Technology_Enhanced_Education_Public_Flagship_Universities_121114_0.pdf.

¹⁷ For preliminary evidence suggesting that advances in online learning might enable institutions of higher education to “bend the cost curve,” see Deming, David J., Claudia Goldin, Lawrence F. Katz, and Noam Yuchtman. “Can Online Learning Bend the Higher Education Cost Curve?” NBER Working Paper 20890 (2015).
<http://www.nber.org/papers/w20890.pdf>.

A second area in which further research is needed is identifying how particular features of online and hybrid instruction impact learning outcomes.¹⁸ As discussed in the prior section, the studies included in this review often conflate online and hybrid courses with different characteristics, making it difficult to tease out the impact of particular characteristics on students' learning outcomes. While a few studies speculate as to the characteristics of the delivery formats that may have driven their results (e.g., degree of interactivity, balance between face-to-face and online material in a hybrid course, etc.), there exists little experimental or empirical evidence in this arena. One promising research design would be to randomly assign sections with different features to examine whether or not the inclusion or omission of those features impacts learning outcomes. This sort of research is necessary to go beyond the question of whether online instruction, writ large, is as effective as that of face-to-face to the more practical question of how to make online instruction more effective.

A third area in need of further study is the effect of online and hybrid instruction in upper-level and humanities courses. Although less common than online and hybrid courses in introductory STEM fields, such courses do exist—often as a result of collaborations among small institutions.¹⁹ The motivation is that these small institutions do not have the faculty expertise or enough interested students to offer every upper-level or humanities course for which there exists student demand. Collaboratively created online or hybrid courses effectively allow these institutions to expand their course catalogs at a reasonable cost. This is a burgeoning and potentially productive use of online instruction without a considerable evidence base.

Finally, researchers should devote more attention to the heterogeneous effects of online instruction and to its ramifications for longer-term student outcomes. As Bowen (2013) points out, online and hybrid instruction are viewed as having the potential to increase postsecondary access and decrease time-to-degree for lower-income and otherwise disadvantaged students, but they may also exacerbate achievement gaps if such students perform less well than their more privileged peers in online and hybrid courses.²⁰ Only

¹⁸ For a brief discussion on the status of the research in this area, see Bell, Bradford S., and Jessica E. Federman. "E-learning in Postsecondary Education." *The Future of Children* 23, no. 1 (2013): 165-185.

¹⁹ Ithaka S+R is currently engaged as an advisor to a Council of Independent Colleges (CIC) initiative entitled the "Consortium for Online Humanities Instruction," which seeks to explore how online learning technologies—particularly in upper-level, humanities courses at liberal arts colleges—can be harnessed to improve student learning outcomes and reduce instructional costs. <http://www.cic.edu/Programs-and-Services/Programs/Online-Humanities/Pages/default.aspx>. <http://sr.ithaka.org/blog-individual/does-online-learning-have-role-liberal-arts-colleges>.

²⁰ See Bowen, William G. "Academia Online: Musings (Some Unconventional)." Stafford Little Lecture, Princeton University, October 14, 2013. <http://ithaka.org/sites/default/files/files/ithaka-stafford-lecture-final.pdf>.

three of the included studies attempt to disentangle effects among students from different backgrounds and with disparate characteristics, whereas the rest focus on average effects over decidedly diverse groups of students. Furthermore, no study goes beyond course-specific outcomes to study the longer-term academic outcomes of students who take online and hybrid courses, such as retention, graduation rate, and time-to-degree. Understanding these heterogeneous and longer-term effects of online instruction is crucial to answering the question posed by Bowen (2013).

Conclusion

Institutions of higher education today are increasingly asked to do more with less, and online and hybrid instruction seem to present a means for colleges and universities to meet their missions more efficiently.²¹ Driven in part by increasing investments in educational technology, online and technology-enhanced learning tools have continued to proliferate in recent years.²² Yet while the potential benefits of online learning have been widely discussed,²³ there is still too little known about the extent to which students have realized these benefits.

The studies reviewed in this report do not thoroughly fill this gap. On one hand, the most methodologically rigorous studies in this review join a growing list of similarly rigorous research finding that students in online and hybrid formats perform about as well as their counterparts in face-to-face sections. On the other hand, there remains a critical need for more rigorous efforts to test the robustness of these effects to various specifications, student groups, and settings.

Moreover, there remains a need for further research on the costs associated with online instruction and the particular features of online instruction that drive their impacts on learning outcomes. It is in extending research to these arenas that one can ultimately assess how online and hybrid delivery formats can be implemented feasibly and most effectively.

²¹ Ithaka S+R recently published a study that finds that rising costs in Virginia's public colleges and universities have been disproportionately borne by students (through increases in tuition) and that net prices in recent years have grown fastest for the lowest-income students—who, in turn, respond most adversely to increasing costs. As a result, initiatives that include experimentation with online and hybrid courses are crucial in efforts to bring down the cost and increase the scale of instruction as well as to improve success rates particularly for those from more disadvantaged socioeconomic backgrounds. For the full study, see Mulhern, Christine, Richard R. Spies, Matthew P. Staiger, and D. Derek Wu. "The Effects of Rising Student Costs in Higher Education: Evidence from Public Institutions in Virginia." Ithaka S+R (2015). http://www.sr.ithaka.org/sites/default/files/reports/SR_Report_Effects_of_Rising_Student_Costs_in_Higher_Education_Virginia_030415.pdf.

²² See Singer, Natasha. "Silicon Valley Turns Its Eye to Education." *The New York Times*, January 11, 2015. <http://www.nytimes.com/2015/01/12/technology/silicon-valley-turns-its-eye-to-education.html>. Blumenstyk, Goldie. "Companies Promise 'Personalized Education.'" *The Chronicle of Higher Education*, September 15, 2014. <http://chronicle.com/article/Companies-Promise/148725/>.

²³ For a stylized model describing one of the potentially broad effects of online learning (in the form of democratizing education), see Acemoglu, Daron, David Laibson, and John A. List. "Equalizing Superstars: The Internet and the Democratization of Education." *American Economic Review: Papers and Proceedings* 104, no. 5 (2014): 523-527.

Description of Studies

Joyce et al. (2014)

At Baruch College (a “large, urban, public university” that is part of the City University of New York), Joyce et al. randomly assign students into “compressed” and traditional formats of an introductory microeconomics course. This course—“Principles of Microeconomics (ECO 1001)” —is required of all students applying to Baruch’s Zicklin School of Business, which enrolls 12,000 undergraduate students (most of whom commute to campus and attend full-time), and fulfills a social science requirement for non-business students. As a result, nearly one thousand students usually enroll in ECO 1001 each fall.

In this study, the authors examine 725 students randomly assigned into four sections of the course, of which two were “compressed” and two were traditional. Students were given an incentive of five extra-credit points on their course average if they chose to participate in the study. The traditional section was offered twice a week for 75 minutes each over a 14-week semester, and the compressed section (supplemented with online material) met once a week for 75 minutes. Two professors taught these four sections (one of each format), and students in each section had access to the same course materials, lecture slides, and pre- and post-lecture quizzes. However, the lecture slides were covered more “selectively and quickly” in the compressed format. The authors evaluate student outcomes based on academic performance on the midterm and final exams (which consisted of the same questions in each section and were all administered in class) and on the final course grade (which also incorporated low-stakes online quizzes on Aplia).

In terms of baseline characteristics in the pooled sample, there are no statistically significant differences between delivery formats on any of the individual characteristics in the initial sample, and only one statistically significant difference (on age) between formats among students who finished the course. While there are some significant differences in baseline characteristics within professor and classroom (particularly in regards to race and prior academic experience), the overall balance is nonetheless favorable and indicates relatively successful randomization.

Across all performance measures in the pooled sample, the authors find that students in the traditional format performed better than students in the compressed format (with these differences being statistically significant for the most part). Incorporating student-level covariates slightly narrows the average differences between formats, but the similarity in coefficients with and without the control variables speaks to the robustness

of the research design. Students in the compressed section scored 3.2 percentage points less on the midterm compared to their peers in the traditional format, but the differences between formats on the final exam were half as large as those on the midterm and no longer statistically significant. The authors suggest that this was a result of students in the compressed section becoming more accustomed to their format as the semester wore on.

The authors also perform analyses under other specifications to ensure that their original results in the pooled sample are as robust as possible. In order to disentangle possible effects associated with heterogeneous professors and classroom sizes, Joyce et al. present estimates separately for each professor/classroom. They find that the differences between the delivery formats were more pronounced in the larger lecture hall, whereas the differences were less substantial and statistically insignificant in the smaller classroom. In fact, in examining within-day student performance on exams, the authors find that students in the compressed section scored more than 5 percentage points less on the combined midterm and final than their peers in the traditional class when the compressed section was delivered in the large lecture hall, but the difference between formats was essentially zero when the compressed class was given in the smaller classroom. Furthermore, they find that high-performing students did equally well regardless of delivery method, whereas the most consistent differences occurred among students in the middle tercile of expected performance.

Finally, the authors observe that the compressed format was not very costly to produce, with advanced testing software and e-textbooks available from publishers at a lower cost than that of a traditional textbook. Furthermore, Joyce et al. mention the potential gains in faculty productivity (measured by faculty compensation per student) and better use of limited classroom space as sources of savings for classes of comparable structure.

Compared to many other studies included in this literature review, the experiment here is one of the most methodologically robust examinations of effect of a hybrid course format on student outcomes. Not only is it the only study with the ability to randomize students into different sections (with a 96 percent participation rate), but its relatively large sample size also increases the precision of its estimates. Additionally, the degree to which the course formats were standardized across sections, combined with the additional analyses and robustness checks that were carried out, confirm the strength of the author's results across various specifications. While there are a few issues that the authors could have more thoroughly addressed, these are relatively minor in the "grand scheme of things."

First, Joyce et al. note that the total post-randomization attrition rate among all students was 9.5%, with evidence showing that there was not selective attrition between formats.

In particular, they find similar results for the midterm exam grades that first included only the students who took the final exam, and then included those students plus those who withdrew from the class after the midterm. However, the authors are not able to determine whether or not there existed selective attrition prior to the midterm exam, which would seem to be the time during which the majority of “leavers” withdraw from the course. Moreover, while each of the sections were randomized in a sufficiently robust manner, students still had to voluntarily agree to be in the study (and were given the incentive of a five percentage point participation bonus). It would be interesting to see whether or not the entire sample of students surveyed across all formats was relatively similar in characteristics to students who took the course in other years. While any year-to-year differences would not impact the internal validity of the study, they may impact its external validity in generalizing to other student bodies.

The authors also mention that they are able to control for instructor-related heterogeneity by having the two participating faculty members teach one course of each format. While this certainly goes a long way in reducing heterogeneity between professors, there still might exist within-professor differences in their abilities to teach each format of the course—particularly if they are teaching one format for the first time. If these differences do indeed exist, then the differences in student outcomes between delivery formats may be driven by both the mode of delivery and the instructor’s ability to teach each course. Finally, the hybrid section in this experiment was essentially a compressed version of the face-to-face section (with the in-class emphasis still on lecturing). To the extent, however, that hybrid courses in other settings might emphasize interactivity over lecturing in their in-class portions, the generalizability of the results in this study may be less powerful.

Kwak et al. (2015)

In this study, Kwak et al. conduct an experiment with a first-year introductory statistics course for business and economics students at the University of Queensland (Australia). This course had traditionally been taught in a face-to-face format, involving thirteen weeks of lectures (with each lecture lasting two hours long and repeated twice a week). In this intervention, the authors replace the face-to-face lectures in the sixth and seventh weeks of the course with a blended format that reduced the face-to-face lecture time from two hours to one hour (with the compressed lecture designed to cover theoretical aspects of a topic). As a replacement for the second hour of the course, the authors offer online material (in six- to eight-minute videos) designed to build on ideas presented in the face-to-face format and provide practical examples and applications of theory covered in the face-to-face lecture.

The lectures for this course traditionally consist of a combination of PowerPoint presentations and Excel demonstrations, and student performance is typically based on six online quizzes and a midterm and final exam. For this particular experiment, the authors compare student performance on the third and fourth online quizzes (given during the two weeks during which the blended format was implemented) to performance on the other quizzes during the semester. They decide that evaluating this design in the middle of the term would give students time to become familiar with various course learning activities and the online quiz requirements.

Kwak et al. first run ordinary least squares (OLS) regressions of students' quiz scores on a binary variable indicating whether or not the delivery format was blended, controlling for various student-level characteristics. They then use a difference-in-differences methodological approach to exploit the causal effect of the blended learning intervention in two ways: 1) comparing the students' performance in 2013 (the year of the experiment) to the performance of prior students who took the class in 2011 and 2012 and 2) comparing the performance within 2013 of students during the intervention to performance during the weeks with face-to-face learning.

In the OLS results, the authors find statistically significant, negative effects of blended learning on student performance, with blended learning reducing quiz scores by 3.4 percentage points (without controlling for student characteristics), and 1.9 percentage points (after incorporating covariates). These covariates include age, nationality, primary language, and achievement level. However, these estimates do not account for omitted cohort- and quiz-specific effects correlated with the blended format that might confound the OLS results. As a result, in utilizing the more robust difference-in-differences identification method, Kwak et al. find that blended learning now had no effect on quiz score. In fact, the blended format was associated with an increase in quiz score by a statistically insignificant 0.16 percentage points, holding constant both cohort- and quiz-fixed effects. The authors also find heterogeneous effects across genders, with the effect of blended learning negative for male students and positive for female students.

The quasi-experimental research design utilized in this study is certainly among one of the more innovative and robust approaches in determining the causal effect of blended learning on student outcomes. In particular, the difference-in-differences research design ensures that the implementation of the blended learning format was as exogenous as possible, meaning that any relative changes in student performance in the course would have mostly been attributed to the inclusion of the blended format. However, there are a few limitations to this study that are worth mentioning.

First, since the blended format was incorporated in the middle of a face-to-face course, it might be the case that the effects of the blended mode may have been confounded by the

face-to-face lectures that preceded it. In other words, the student outcomes that the authors observe to be associated with the blended format may have been partially driven by the students' prior exposure to face-to-face courses. This may then impact the external validity of this study's results, particularly since most blended formats remain fully blended throughout the balance of a term. Indeed, the course structure utilized by Kwak et al. is relatively "non-traditional," as it would be rare for instructors outside of an experimental setting to devote two weeks in the middle of a face-to-face section to a blended format.

Finally, in evaluating the direct effects of the blended format on student outcomes, the authors rely on quiz grades attained in the weeks that the blended mode of delivery was implemented. This, however, does not serve as a holistic indicator of a student's performance, especially since the effect of a course delivery format may not be entirely robust until observed over the course of an entire term. While the authors also compare final course grades across years (between fully face-to-face sections in previous years to the experimental section they observed), any differences may not be entirely generalizable to external settings, since the blended format only comprised a small fraction of the experimental course in the year it was offered.

Olitsky & Cosgrove (2014)

In this study, Olitsky and Cosgrove compare learning outcomes for 318 students—the majority of whom were sophomores and business or pre-business students—enrolled in blended and face-to-face versions of Principles of Microeconomics and Macroeconomics courses during the 2011-12 academic year. The microeconomics instructor taught one blended section in Fall 2011 and two face-to-face sections in Spring 2012, whereas the macroeconomics instructor taught one blended and face-to-face section each in Fall 2011 and two face-to-face sections in Spring 2012. Both faculty members completed a faculty development course in Summer 2011 that imparted best practices in blended learning. The blended course substituted online instruction (constituting "online lectures, article analyses, discussion board assignments, and group wiki assignments") for one third of the semester's class periods, and the instructors used identical textbooks, course and homework management websites, assignments, and exams within and across delivery formats.

Student outcomes are evaluated via several measures designed to test understanding of a single learning objective (analyzing opportunity costs to determine the most "efficient specialization of production") during the few weeks of the course when the delivery formats were most similar to each other. These measures include an overall exam on this objective, an online homework assignment, and short-answer questions. General learning outcomes over the course of the entire semester are also assessed. The authors

are also able to match each student's assessment results to university transcript and demographic information, finding that students in the blended section tended to have fewer cumulative credits, were less likely to have taken principles of economics before, and were less likely to be first-year or non-international students. The summary statistics also suggest that students in the face-to-face sections performed significantly better on nearly all of the student outcome measures.

Olitsky and Cosgrove first estimate an OLS regression that used a binary variable for blended status as the primary independent variable and controls for individual characteristics (including academic background, prior experience with economics, and race/gender) presumably correlated with the delivery format and the outcomes of interest. They find that, after incorporating student-level covariates, the effect of blended learning on outcomes is statistically insignificant. However, the effect of blending on online homework is only marginally insignificant at the 95% confidence level, with the blended format associated with a decrease in the homework grade of approximately four points.

Given that the OLS specifications cannot fully control for bias stemming from non-random selection into each delivery format, the authors utilize propensity score matching to try and tease out the causal relationship between blended coursework and learning outcomes. At the same time, they emphasize that selection bias in their study may not be as pronounced as in other studies given that students did not know the delivery format of their course until the first day of class (but before the add/drop deadline). Under the propensity score matching methodology, the authors account for a set of covariates (the same ones as those employed in the OLS regression) that predict the likelihood of a student selecting into a given delivery format and then estimate a propensity score based on this probability.

Olitsky and Cosgrove find via the propensity score results that the blended format once again produced generally statistically insignificant effects on student outcomes, although the magnitudes of these average treatment effects are larger than those generated under the OLS specifications. For example, blended coursework was associated with a decrease in total exam score of 0.65 points in the OLS regression, whereas the estimate under propensity score matching was a decrease of 5.5 points. In analyzing the average treatment effect on those in the treatment group, the results once again suggest no significant effect of blending on learning outcomes, although the lack of statistical significance may be due to the small sample size. However, the effect of blending was consistently negative for the online homework assignment (with decreases between 1.3 and 4 points on homework grades across various specifications), although it disappeared when students took the exam. The authors also estimate quantile regressions to assess

whether the blended format had differential effects across the spectrum of outcomes and find only small differences across the distribution for each outcome.

Among the strengths of this study, the level of sophistication utilized in its analyses goes a long way in ascertaining the causal effect of blended learning from a non-randomized intervention. In particular, the overall consistency in results between the various propensity score matching specifications provides confidence in the general result that blended learning produces no significant effect on student learning outcomes. The authors also do a noteworthy job of evaluating and quantifying sample selection bias, finding that students do indeed select in and out of blended courses (as they do with online classes, as prior research has shown). Olitsky and Cosgrove are also very explicit in demarcating the differences between the blended courses and the face-to-face courses that served as “control” groups and standardizing course materials and instruction. They go beyond other studies in ensuring not only that blended and face-to-face sections were taught by the same instructor, but also that these instructors had somewhat comparable levels of comfort and skill in teaching both outcomes (via the faculty development course in blended learning given prior to the school year).

Among the shortcomings associated with this study, the blended and face-to-face formats were evaluated in different semesters and not side-by-side. While the authors control for some observable covariates (including experience with prior economics courses) that may have varied depending on when the delivery format was offered and were correlated with learning outcomes, there might be other unobserved factors (such as changes in institutional policies) varying over time that may have confounded the results. Furthermore, while the authors employ demographic and academic background control variables in their OLS and propensity score specifications, they do not include any covariates associated with students’ socioeconomic status (such as parental education, family income, etc.). While one could argue that these variables may be collinear with other variables included (such as GPA and race), they are sufficiently independent that omitting some sort of direct proxy for socioeconomic status likely biases not only the OLS results but also the prediction of the propensity scores associated with selection into the blended format.

In addition, the statistical insignificance of the authors’ propensity score results may have been as much a product of the small sample size (and thus large standard errors) as they are a reflection of the causal relationship between delivery format and learning outcomes. The lack of a large sample size is particularly important under propensity score matching in order to reliably predict the selection of a treatment across as broad a sample of students as possible (with preferably overlapping characteristics between students in the treatment and control groups). Finally, as the authors also acknowledge,

it is unclear the extent to which the results in this study are externally valid and generalize to other settings and populations.

Burns et al. (2013)

In this study at a “Midwestern land grant university,” Burns et al. create online and hybrid versions of “Computers and Information Systems (IS100),” a required course for all business majors. During the first round of implementation, two delivery modes—online and face-to-face—were simultaneously employed for four consecutive semesters beginning in Fall 2010. In the second round of implementation, a third hybrid delivery mode—promoted by the National Center for Academic Transformation (NCAT) redesign initiative—was instituted in Fall 2011, which replaced some, but not all, of the in-class meetings with online, interactive learning activities. In addition to measuring the effect of the IS100 delivery mode on the immediate performance of students in their IS100 course, the authors attempt to evaluate longer-term outcomes by examining the subsequent performances of IS100 students in the more advanced “Concepts and Applications (IS200),” also a required course for business majors with IS100 as a prerequisite.

The face-to-face sections met twice weekly, and independent workdays, recorded lectures, instructor office hours, and the identical learning management system were offered to students in all sections regardless of delivery method. For the face-to-face and hybrid sections of IS100, exams were held during regular class meetings and proctored by the instructor, whereas exams for the online sections were allotted the same amount of time but were not proctored by the instructor. Between Fall 2010 and Fall 2012 (the two-year time frame of this study), there were a total of 382 student observations, with 109 enrolling in the face-to-face section, 144 in the online section, and 129 in the hybrid section. However, only 233 students of the original 382 had prior academic achievement records (i.e. GPAs), which are crucial to control for given their importance in predicting student success. Of these 233 students, 130 students that completed an IS100 section offered by the authors later enrolled in IS200.

For each semester of IS100, students self-selected the delivery mode. As a result, the authors find that students self-selecting the face-to-face and online delivery modes had greater prior academic achievements than students self-selecting the hybrid mode, whereas the students self-selecting the face-to-face delivery mode were younger and less likely to be Pell Grant-eligible than those who chose online or hybrid delivery. Finally, males were more likely to enroll in hybrid sections, while students with nearby residences were more likely to enroll in face-to-face sections.

The authors use a set of ordered probit regression models to determine the effect of a class delivery mode on a categorical achievement level conditional on observed demographic and academic background characteristics of each student (in an attempt to control for self-selection effects in learning outcomes). They find that students who took IS100 in the face-to-face version had statistically better learning outcomes than those in the other delivery formats. Nevertheless, the authors observe that students who took online or hybrid version of IS100 actually outperformed those students in IS200 who took the face-to-face version of IS100. They also note that the delivery mode of IS200 was not a significant predictor of student performance in that class, and that prior GPA was a statistically significant covariate in all specifications.

Although this study goes further than many others in attempting to assess the effect of online learning on longer-term performance on a follow-up course, the methodology used to examine this particular question may suffer from a serious flaw. In particular, whereas IS100 attracted students from various majors in addition to business students who took it to fulfill their major requirements, IS200 was taken primarily by business majors. It is not a surprise, then, that only 56% of all students in the sample who completed IS100 enrolled in IS200, with the remaining 44% likely consisting primarily of students from non-business majors. As a result, the seemingly interesting dichotomy in the effect of the IS100 delivery format between student performance on IS100 and the subsequent IS200 course may not be so much a product of the hybrid and online formats themselves as it is an effect of changing student samples between these courses. It would be interesting to note, for example, whether or not the online and hybrid formats for IS100 tended to have larger concentrations of business majors who would later enroll in IS200. This particular issue is, in some ways, a microcosm of the larger threat of sample selection bias, which originates from the students self-selecting into their initial IS100 section. While Burns et al. attempt to address this possible threat to internal validity by controlling for student-level characteristics that appear to be correlated with both selection of the delivery format and their eventual performance, it is unclear to what extent doing so adequately removes bias otherwise existent in their specifications.

Furthermore, the authors note that the instructors in IS100 had far more control of their delivery method than those in IS200. There are two possible methodological issues that might arise from this observation. First, in the absence of, at minimum, instructor- or class-fixed effects, it becomes even harder to assess the controlled effect of the IS100 delivery method on student performance in an IS200 section that was taught in a sufficiently independent manner from IS100. Second, as noted by Burns et al., the instructors' more personal interactions with students in IS100 may have impacted the leniency of the grading standards in IS100 relative to those in IS200.

Given the fact that the hybrid sections were only instituted in 2011 (a full year after the online and face-to-face sections were introduced), the authors should also have controlled for factors varying between 2010 and 2011 (i.e., through year-fixed effects) that may have impacted both the delivery format and student performance. Moreover, it is concerning that the exams in the face-to-face and hybrid sections of IS100 were proctored, whereas exams in the online sections were not proctored. The authors do not mention what steps, if any, they took to ensure that students in their research design were administered exams on a “level playing field” regardless of the delivery method they chose.

Verhoeven & Rudchenko (2013)

Verhoeven and Rudchenko conduct this study at a “large public university” where the majority of students work and commute to campus. They examine an undergraduate Principles of Microeconomics course, in which 51 students enrolled in a hybrid section and 24 students enrolled in a face-to-face section. Both sections were taught during the same sixteen-week term by the same professor and had identical course materials, exams, and a “Desire2Learn” course management system. For this class, the hybrid section met for 1.25 hours once a week (with class time exclusively devoted to PowerPoint lectures) and the face-to-face section for 2.75 hours once a week (with class time devoted to the same set of PowerPoint lectures in addition to students working on practice problems in an interactive manner).

The authors find that 71% of the students in the hybrid section and 79% in the face-to-face section completed the course, although these rates are not significantly different from each other. As a result, the number of student observations for each course drops to 36 and 19 for the hybrid and face-to-face classes, respectively. Given that students in the hybrid section had a higher average GPA than their counterparts in the face-to-face class, Verhoeven and Rudchenko control for GPA in an ordinary least squares (OLS) regression of composite test score on the delivery method and find that hybrid delivery was associated with an estimated test score that was 4.8 percentage points lower than that under face-to-face delivery. The authors explain that this outcome is likely a result of the hybrid section having “no required learning activities to compensate them for the ongoing in-class practice working analytical problems afforded the face-to-face students.” In fact, the use of online resources for students in the hybrid section was restricted to non-interactive content.

While the authors incorporate several important facets of a controlled experiment into their study (including side-by-side analysis and standardizing course materials, assessments, and instructor), they are unable to conduct a fully randomized control experiment (RCT), as the students self-selected into each delivery format. Furthermore,

the sample size in this study is quite small—even by standard traditionally associated with RCTs. This may then explain the large standard error—and thus lack of statistical significance—for the coefficient on hybrid delivery. Nonetheless, the small sample size is made even more concerning by the fact that 29% of students in the hybrid section and 21% of students in the face-to-face section failed to complete the course and, therefore, were not included in the final analysis of student outcomes. The focus on student outcomes upon completion and the failure to analyze the factors behind student attrition are not issues unique to this particular study, but the magnitudes of the attrition rates in this experiment are quite concerning. Consequently, the observed association between test score and the hybrid format may have been just as much a factor of the changing student sample as it is of the delivery format.

Moreover, Verhoeven and Rudchenko do not do enough to control for factors correlated with delivery format and course grade that may, when omitted, bias the results they observed. While controlling for prior GPA is an important step, they should also have looked towards other covariates such as GPA in concurrent courses, age, socioeconomic status, and hours worked during the semester. In addition, the authors note that there existed “noise inherent in the calculation of the composite test score,” meaning that their primary dependent variable of interest may have been imprecisely estimated to begin with.

Finally, the distinction in design between the hybrid and face-to-face formats in this study is different from the distinction employed in other cases. In particular, whereas the hybrid format in other settings may emphasize more student-to-student and student-to-teacher interactions in lieu of lectures during the in-class section, the hybrid format here was devoted entirely to covering the same set of lecture slides as the face-to-face section (which then used the extra time to facilitate student interaction). In other words, the face-to-face section in this study adopted some of the aspects commonly unique to the hybrid format, while the hybrid section itself became virtually a condensed version of a face-to-face lecture. As a result, it is very difficult, if not impossible, to generalize the results in this experiment to other settings where the hybrid and face-to-face formats may be structured differently. In fact, the negative results found in this study between the hybrid format and student test scores could very well speak to the positive effect of in-class interaction (which is traditionally a hallmark of the hybrid format).

Xu & Jaggars (2014)

In this study, Xu and Jaggars examine the performance gap between online and face-to-face courses and how this gap varies across subgroups of students and academic subjects. They use an administrative dataset covering enrollment in nearly 500,000 online and face-to-face courses taken by more than 40,000 degree-seeking students who

initially enrolled in 34 community or technical colleges in Washington State during Fall 2004. The dataset contains a rich variety of information for each student on demographics, socioeconomic status, academic background, and wage records. Xu and Jaggars follow each student for five full academic years through Spring 2009 and focus primarily on assessing the impact of delivery format on course persistence and grade.

In their regression analyses, the authors first use an ordinary least squares (OLS) model that regresses some indicator of course performance (i.e., persistence or grade) on a binary independent variable indicating course delivery format and a set of student-level covariates and term- and subject-fixed effects. To deal with variation in grading standards within a particular subject area, Xu and Jaggars convert course grade into a standardized z-score that represented a student's performance relative to that of other students in standard deviation units. The authors then incorporate individual-fixed effects to account for unobserved factors that may affect an individual student's likelihood of choosing online coursework, as well as a covariate for the average persistence rate of a given course to deal with course-level variation in instructional quality that might be correlated with student outcomes. Finally, Xu and Jaggars conduct a series of additional robustness checks to examine whether individual differences that varied across time may have biased their initial results.

Approximately 10% of all course enrollments were taken online, with the online format being most popular in humanities, social science, education, and computer science courses. In terms of summary statistics, the authors find a noticeable gap in persistence rates between online (91.17%) and face-to-face (94.45%) courses, and in average grades on a 4.0 scale between formats (2.77 for online versus 2.97 for face-to-face). In the OLS specifications, the authors find that the effects of the online format on course persistence and standardized course grade are consistently significant and negative across all models. These estimates are even larger after incorporating individual-fixed effects and, at minimum, remain comparable after considering previous online learning experiences and current working hours.

Xu and Jaggars then proceed to examine any heterogeneous effects of delivery method on different subgroups of students and subjects. While they find negative effects of online learning across every subgroup, men had stronger negative estimates than women for both course persistence and course grade, and black students were twice as likely as Asian students to be negatively affected by an online course (in terms of grade). Furthermore, younger students had stronger negative coefficients for online learning than older students, although these estimates were statistically significant in both cases. The authors also find that students with a stronger academic background had narrower gaps in online performance, whereas students with weaker skills had wider gaps (compared with students in the face-to-face courses). Finally, Xu and Jaggars observe

negative coefficients for online learning across every subject area, although there are variations in statistical significance (with education, mass communication, and health and physical education having insignificant estimates) and magnitude (with weaker coefficients in natural science and stronger estimates in English). Furthermore, these performance gaps become wider when students took subjects that enrolled more online, at-risk peers.

This study was executed in a very complete and ambitious manner, and Xu and Jaggars are to be commended for rigorously analyzing the impacts of online courses in community colleges, an important but often overlooked research avenue. Especially impressive is the fact that they are able to exploit student-level outcomes across such a wide range of course enrollments, in addition to following individual students longitudinally to assess how delivery formats are associated with longer-term outcomes like retention. Furthermore, the authors' analyses of heterogeneous effects across different student groups are particularly strong and speak to the value of disentangling average effects into more precise relationships with actionable implications.

Nevertheless, it is worth pointing out a few methodological issues arising from this non-experimental study. First, although Xu and Jaggars control for a wide range of covariates and fixed effects in order to mitigate the potential biases from the non-randomization of students into delivery formats, it is unclear whether or not they control for enough variables to completely eliminate omitted variable bias. For example, Xu and Jaggars control for term- and subject-fixed effects to account for unobserved heterogeneity between academic terms and course subjects—however, they did not say why they choose not to control for institutional-fixed effects. While one might argue that there is less variation in the policies implemented and the institutional backgrounds among two-year colleges than among four-year schools, there still likely remain unobserved differences between community colleges that would be worth addressing. Furthermore, subject-fixed effects might not account for within-subject differences across courses stemming from variations in difficulty, instructional quality, and course materials and platforms. The authors could have also incorporated a time trend within courses to address not only unobserved heterogeneity between course structures but also term-to-term differences within a course that would otherwise not be captured (e.g., changes in instructors within a course).

Finally, one major methodological shortcoming in this study pertains to the authors' inability to distinguish between modes of delivery within online courses. In other words, hybrid and fully online courses were all included within the "online" classification, with Xu and Jaggars giving no indication about what the online courses encompassed and what sorts of variations existed between these courses. This is particularly concerning, given that the large size and diversity of the institutions and courses sampled almost

certainly result in substantial differences in formats between online courses. Were the observed effects of online courses driven primarily by formats that were fully online, or were they driven more by technology-enabled courses with more substantive face-to-face formats? Without these more precise distinctions, it becomes difficult to conclude to what extent Xu and Jaggars' findings are applicable to certain types of technology-enhanced courses.

Carmichael et al. (2014)

In this study, Carmichael et al. examine learning outcomes associated with online, hybrid, and face-to-face (“on-ground”) versions of a three-credit interdisciplinary capstone course entitled “Writing Across Disciplines.” The authors taught this class at the University of North Dakota, a comprehensive research university that enrolls more than 15,000 students across its undergraduate, graduate, and professional programs. This course was developed in the Humanities and Integrated Studies unit, and is offered as an “upper-level, intense writing experience” requirement for seniors who have already completed basic required writing courses. Students enrolled in this course came from a wide range of majors and varied in their prior experience in student-centered interdisciplinary classes.

All three sections of this course were taught by the same professor and used the same texts and assignment prompts. In the face-to-face version, a small group of seniors would come together for in-class discussions and activities, which included writing assignments, presentations, and student-to-student conferences and peer review sessions. In contrast, the delivery method in the online version was completely asynchronous, with student interaction solely facilitated in the Blackboard online course forum. In the hybrid version, students met face-to-face for three-week blocks, with additional online writing assignments alternated with online work for three-week blocks. The online version of the course was offered multiple times, whereas the face-to-face and hybrid versions were each offered once. All courses had the same number of students, along with a predominance of similar majors across all versions.

In this pilot study, formal writing assignments were collected at the beginning and end of the semester, with assessment based on rubrics for Writing, Critical Thinking, and Integrative Learning (which ranked students on a scale from 0 to 4 for each criteria). Reviewers of the assignments did not have any knowledge of the course version, and results were averaged across each category. Across all delivery methods, average scores for the first paper were lower than expected (with averages ranging from 2.04 with a standard error of 0.40 to 2.26 with a standard error of 0.30). Furthermore, comparisons of the first and last papers demonstrate that average rankings increased in all three rubric areas across all delivery methods. Nevertheless, the improvement is statistically

significant at the 95% confidence level only for one of the online sections and for the hybrid section. Interestingly, rankings for the face-to-face version tend to be the lowest of all versions, with the average rankings for the hybrid section more comparable to those for the online course than the face-to-face section.

There are several features of this study for which the authors should be commended. First, they tackle a subject area that is admittedly more difficult to assess (given the subjective nature of evaluating writing, compared to the more objective evaluation of, say, grading statistics exams) and has thus been relatively marginalized in the literature surrounding the impact of various delivery formats. Given that more and more courses with substantial writing components are experimenting with online and hybrid formats, it remains imperative that more studies be carried out to rigorously assess their effectiveness. Carmichael et al. take an important step in this endeavor. Furthermore, the utilization of pre- and post-tests in this research design goes a long way in ensuring that the results try and reflect the change in outcomes associated with a specific section (rather than purely an outcome at the end of a section, which could be tainted by exogenous factors that may—to an extent in this case—be captured by the pre-test outcome). Finally, the authors make substantial efforts to standardize instructors, course materials, and class sizes across delivery formats.

However, there still remain some limitations to the explanatory power embedded in these results. The self-selection of students into delivery formats raises concerns about the comparability between different samples of students among each section. While the implementation of a diagnostic evaluation at the beginning of the semester helps the authors get a sense of what this sample “looks” like, it would have been even more rigorous to go beyond the purely descriptive analysis and control for existing characteristics correlated with both selection of the delivery format and course performance in a regression analysis. Furthermore, there may also be student characteristics that vary between sections during the semester that, when omitted, could confound the results. These include indicators for workload in concurrent courses and part- or full-time work outside of classes.

It would also have been helpful for the authors to present a broader description of the distribution of outcomes in each delivery format rather than merely report and analyze the average. Was there a significant difference between the median and mean for each section, or was the distribution of scores relatively constant? Were there any significant outliers, and, if so, did these skew the average scores reported in any way? Because the outcomes are recorded categorically (as integers between 0 and 4) rather than continuously (e.g., as percentages between 0 and 100), it would not have been overly complicated to show this. Carmichael et al. also acknowledge that the hybrid format employed in this experiment had never been taught before, which means that there may

have been “unevenness due to instructor issues” that may not have been controlled for. Consequently, the results may differ in future semesters—and thus be more generalizable to external settings—when the hybrid section is more systematically taught.

Fish & Kang (2014)

Fish and Kang compare outcome data obtained from student enrollees of an undergraduate, upper-division stress management course taught at a “large, public university of the West Coast of the United States.” Because the course fulfills a general education requirement, it consisted of students from all majors across the university. One section was taught in a face-to-face format, and the other was offered completely online. Both courses were taught over a 10-week period and featured the same instructor and consisted of identical assignments.

The sample sizes for each section are fairly similar—56 students in the online section and 63 in the face-to-face section. In terms of additional similarities between the two course formats, the lectures, exams, and course requirements consisted of the same content and instructions. However, online students viewed and/or listened to recorded lectures, whereas face-to-face students listened to live lectures offered twice a week (with each session lasting for approximately 100 minutes). Face-to-face students also had opportunities to discuss questions in small groups and somewhat modify the substance of the lecture through in-class questions. Because attendance was counted as part of a student’s grade in the face-to-face format, students in that section were incentivized to attend class. Finally, exams in the face-to-face section were administered in a proctored, in-class environment, while exams in the online version were offered via Blackboard in an un-proctored environment with various limitations (which included timing, randomization of questions, and inflexibility with regard to question order).

Using a t-test, the authors find that no significant differences in exam scores between the delivery formats when all three exams given throughout the term were examined together. However, in analyzing exams one-by-one, Fish and Kang find that there was a statistically significant difference among scores for the third exam—in favor of the online format. Moreover, older students (particularly on the second exam) scored lower than younger students, and Latino students scored lower than Caucasian students. It may be useful to note that the exam scores were one of several dependent variables that the authors analyzed in their study, although the others were largely self-reported (on the part of students) and significantly less objective.

To their credit, Fish and Kang analyze the impact of various delivery formats in a non-traditional course that may present difficulties in evaluating student outcomes that are somewhat different from those well documented in the literature on more “traditional”

courses. However, there exist some methodological flaws in their research design that are worth pointing out. First, the authors fail to contend with two forms of sample selection bias. On a broader level, they acknowledge that the students who took the stress management course during the semester of interest did so voluntarily and under the knowledge that they were participating in a study. Therefore, there may have been a non-random group of students that chose not to take the course as a result of not wishing to participate in a study. Furthermore, the students themselves self-selected into each delivery format, as the authors are unable to randomize students into particular sections. Moreover, Fish and Kang admit that the sample sizes from which they draw their results are quite small.

While it is understandable that the sources of selection bias might come from factors outside of the researchers' direct control (e.g., having to attain full Institutional Review Board approval), the authors could have done more to control for variables correlated with student outcomes and delivery format. For example, the authors observe that students in the online course were slightly older than their counterparts in the face-to-face section. This distinction—when unaccounted for—may be particularly dangerous in this setting, as age would appear to be particularly correlated with the subject matter of this course (stress management). While the authors do present summary statistics associated on how exam grades vary across students with different characteristics, it would have been far more valuable to incorporate these characteristics as covariates in a multivariate regression of student outcomes on delivery format.

Finally, the online and face-to-face sections are not completely standardized, which means that the measures of student performance associated with each section may have been driven not so much by the format of the course delivery as by additional factors unique to each section. For example, exams in the face-to-face section were administered in a proctored, in-class environment, whereas exams in the online section were given online in an un-proctored setting. While the authors try to ensure that these environments are as standardized as possible, there may still have been some significant differences that impacted the validity of the evaluation. Furthermore, participation in the face-to-face course was counted as part of the course grade, while Fish and Kang do not give evidence that a similar grading structure was implemented in the online section. As a result, students in the face-to-face course may have been more incentivized to “attend” lectures, thereby inflating the exposure they had to instruction (relative to the students in the online course) that thus provided potentially positive effects for their performance. The authors also state that the online learning model used in this course was relatively “bare-bones,” and without much “multimedia, discussion boards, or videos.” This may provide limitations to the external validity of this study, particularly in settings where the online courses are developed with more sophistication.

Jones & Long (2013)

Jones and Long gather final course grades from a mathematics course entitled “Quantitative Business Analysis I (QBA I)” at a “small, open-enrollment rural Appalachian college” over the course of ten semesters. In each semester, one section of the course was offered in an online format while the other was offered in an on-site format. This course is required for all Business majors, who usually take the course during their freshman year. While one instructor taught the online section consistently across all semesters, four different instructors who employed their own grading and assessment systems taught the on-site section.

In terms of descriptive statistics, 267 students took the on-site version of the course and 178 students took the online version, with the mean and median grades higher for the on-site than for the online sections (with this difference being statistically significant at the 95% confidence level). In both sections, there is a negative skew in the distribution of scores (which appears to be significantly non-normal) and similar levels of variance in final grades. However, on-site students had a larger range of scores compared to their online peers, as well as a greater spread of grades in the middle 50% of the data.

However, a semester-by-semester analysis of the grades reveals that the instructor teaching the on-site section during the first three semesters may have been more generous with regard to grading than subsequent instructors. At the same time, grades in the online section may have been lower in the first few semesters due to some adjustment on the part of students to a new method of learning. As a result, the authors conduct another analysis that omits the first three semesters from their dataset and find that, this time, no significant difference was observed between the mean scores for on-site and online students.

While this study examines multiple semesters’ worth of data on the efficacy of online and hybrid courses, it is not without its limitations. First, the authors do not make an attempt to randomize students across online and on-site formats, which means that students self-select into their delivery format of choice. While the inability to conduct a fully randomized controlled experiment is understandable (given how complicated it can be to put together the experiment), it would have been desirable for the authors to control for observed and unobserved differences between students and classroom environments (from prior academic achievement and socioeconomic status to instructor quality) in each of the delivery formats. Not doing so results in myriad of threats to internal validity, including omitted variable and sample selection bias. As a result, even while Jones and Long may have set out to execute an observational analysis in this study rather than a stronger regression analysis, it is very difficult for one to conclude that the relationship

observed in this study between delivery format and student outcomes is robust, let alone causal.

One of the biggest sources of bias in this study likely comes from the observation that the online course was taught by one instructor over this time period, whereas the on-site section was taught by four different instructors—none of whom was the one who taught the online section. As a result, instructional quality, course standards, course materials, and assessment mechanisms may have varied greatly both between formats and across semesters within the on-site format—with these factors likely confounding the study’s results. In fact, Jones and Long find that omitting the observations from the first three semesters of the study (when the on-site instructor was known to have had more lenient grading standards) significantly changes the overall results. While it is reassuring to see the authors accounting for these “anomalies” in the analysis, they also raise questions about the possible existence of other instructor-related sources of bias when the inconsistencies are not as obvious.

Jorczak & Dupuis (2014)

In this quasi-experimental study, Jorczak and Dupuis examine student outcomes associated with face-to-face and online sections of an introductory psychology course at a “medium-size Midwestern public university.” This course is required for psychology majors and was divided into one in-class section that met three times a week for one hour each, and two online sections that were delivered via a course management system. The in-class meeting enrolled 35 students, whereas the two online sections enrolled a total of 69 students. Both delivery formats covered identical material that followed the same sequence and schedule and were taught by the same instructor. However, whereas lectures were presented live to students (who in turn would participate face-to-face in small-group discussions) in the in-class format, online students received brief text-based “lectures” and interacted with other students in an asynchronous manner. Students were not randomly assigned to sections in this research design.

Students in all sections took the same two 50-item exams, each of which were timed and employed multiple-choice items that were scored for accuracy. Regardless of the course format, these were delivered via the university’s online learning management system. The authors find that students in the online section scored an average of 74% on the exams compared to 67% for the in-class students, with this difference statistically significant. In particular, an average student who participated in online discussion was predicted to score 25 percentile points higher (a “moderate to large” impact) on course exams than the comparable in-class student. Jorczak and Dupuis also find a moderate association between online discussion participation and exam performance (with a statistically significant correlation of 0.36), although the association between exam

scores and course grade is relatively weak. Although the authors are unable to control for omitted variable bias, they did observe that students' grades on the first quiz score—a proxy of their knowledge and test-taking skills prior to the course—are not very different across the delivery methods.

Jorczak and Dupuis do an admirable job of normalizing course structures and materials across sections. Not only were the same course materials delivered over the same sequence and taught by the same instructor in both sections, but the exams were also identical in both substance and delivery. This last similarity is particularly important and unique, as most studies that examine outcomes from online and face-to-face sections administer exams in the delivery format that correspond with each section (which often result in problematic differences between testing environments). However, the author's research methodology also suffers from a few shortcomings, which they nevertheless make an effort to spell out.

The authors admit that students were not randomly assigned between sections, which thereby damages the ability to draw causal connections between the course delivery format and student performance. Jorczak and Dupuis, however, make an attempt to measure the initial comparability between students in each section by comparing student outcomes on the first quiz of the semester and find that there is no statistically significant difference between students in each section. However, even if one is to make the generous assumption that this quiz served as a robust proxy for prior student achievement, there could have been other non-academic factors existent prior to taking this course (e.g., family income, race, parental education), as well as factors that changed over the semester (e.g., concurrent course load, part- and full-time work status) connected to both the selection of the delivery format and eventual performance. By not controlling for these factors in their study, the authors expose their study to the threat of omitted variable bias.

The sample size for this study is also not large (with a total of 104 observations), which means that the authors would have had to be even more attentive to sample selection and omitted variable bias to ensure as robust a research design as possible. Moreover, Jorczak and Dupuis identify one potentially confounding difference between the structures of each section (which were otherwise relatively well standardized). In particular, students in the online section were graded for participation in discussion, whereas students in the face-to-face class were not assigned points for participation. To the extent that students in the face-to-face section received a relatively lower incentive to participate relative to their online peers, this difference may have biased the degree to which student outcomes were associated with delivery format (especially if class participation is correlated with exam scores). Finally, given that the authors' main dependent variable of interest (exam score) had a “surprisingly weak” association with

course grade, there remain concerns as to whether or not exam performance serves as a suitable proxy for their student performance in the course as a whole.

Metzgar (2014)

This study takes place at a “large Southern public university in the United States,” where a hybrid approach was implemented across 3 sections (with each containing 80 students) of a junior-level Managerial Economics course required of all business majors. The hybrid format used the “MyEconLab” online platform from Pearson, which was combined with the purchase of an e-textbook for the course. Moreover, class time for the hybrid section was reduced from two 75-minute periods per week to one 75-minute period and was focused on student questions from the homework and quizzes for the first 10 to 15 minutes of the class. The rest of the class time was given to reviewing concepts.

To analyze the effect of the hybrid approach on student outcomes, Metzgar used in-class clicker questions in the hybrid section identical to those from the previous semester, and the same exam questions from previous semesters were also used in the hybrid sections. While the author taught all sections of the hybrid course as well as the traditional courses in previous semesters (used as the comparison group), the analysis in this study is less of a controlled experiment than it is more of an observational analysis of changes in student outcomes associated with the hybrid format.

In general, while students in previous classes would often get 70-80% of the clicker questions correct, students in the hybrid section would only score 30-60% (with significantly more variance). The differences in scores were smaller on the exams, with students in the hybrid sections scoring between 60-70% (compared to 70-80% in previous semesters on identical questions). At the same time, the hybrid approach resulted in an increase in the amount of total hours devoted by the instructor to the class (with most of the increased workload spent on computer-related tasks such as fixing glitches). In the end, Metzgar concludes that the unsatisfactory effects of the hybrid approach may have been due to the complexity of the course material (which may be better suited to a face-to-face format).

While the instructor employs identical clicker and exam questions in the face-to-face and hybrid versions of the course, each delivery format was offered in different semesters. As a result of this “before-and-after” research design, the results observed by the author likely do not allow for a causal interpretation, particularly since they do not account for factors changing over time that may be correlated with each section type and with student performance. Were there institutional policies, for example, implemented after the face-to-face course was offered that impacted how the hybrid course was delivered?

To what extent were course materials and the course structure, other than the substance of the in-class clicker questions and exams (which were identical across semesters), uniform in each semester? Were the class sizes in the face-to-face and hybrid sections similar, and were the in-class portions of each section offered in classrooms of similar infrastructure? It would have been useful to have more information on possible changes in these class- and institutional-related characteristics in determining the degree to which the observed changes in student outcomes were directly driven by the change in course delivery format.

Furthermore, Metzgar acknowledges that “adjustments were not made for potential differences in student characteristics.” This is particularly concerning given that this study not only failed to randomize students to each section, but it also did not capture how the characteristics of students in each section may have changed from semester to semester. For example, how did the academic backgrounds and family incomes of the students in the hybrid sections compare to those in the face-to-face format? Were the hybrid students older on average than the students who took the face-to-face section? Did they have to take heavier concurrent course loads? If these differences did in fact exist between the delivery formats, then they may very well threaten the internal validity of the results in the study.

Tanyel & Griffin (2014)

Tanyel and Griffin, at a “southeastern regional university” that enrolls approximately 5,000 mostly full-time students, carry out a ten-year longitudinal analysis of student outcomes between Spring 2002 and Spring 2011 in online and face-to-face courses. Specifically, they compare the grades and persistence rates of students who took online and face-to-face sections of courses taught by the same instructor within the same semester (with 15 of the 19 semesters in this time period containing classes that fulfilled this criteria). However, the authors cannot distinguish between different online modes of delivery (with hybrid, blended, and fully online courses all falling under the “online” umbrella), nor could they determine the comparability of course materials and assignments between formats. They used a dataset that includes information about all students enrolled in undergraduate, non-nursing courses (as nursing courses had stricter criteria for entry), including such characteristics as major, prior GPA, age, and course performance.

Over the ten-year period, there were 38 different instructors who taught both online and face-to-face versions of the same course during the same semester. These amounted to a total of 81 different courses (with 66 from the College of Arts and Sciences, 11 from the College of Business and Economics, and 4 from the School of Education) divided into 132 face-to-face and 94 online sections. There were 5,621 students in the total sample (with

3,355 students in the face-to-face sections and 2,266 students in the online sections). Nearly half (49%) of students in the online course were over the age of 25 (compared to only 26% of students in the face-to-face versions), and 20% of online students were taking upper-division courses (versus 12% of face-to-face students). However, there are no significant differences in the proportion of students with prior GPAs over 2.5 between the face-to-face and online sections. Even after condensing the sample to include only those students who finished and received a grade for the course, Tanyel and Griffin find that the baseline characteristics for the new sample remain largely unchanged from those for the original sample.

There are, however, differences in student success between face-to-face and online sections. In particular, 30% of students in the online classes either failed or withdrew from the class—compared to 18% of students in the face-to-face classes (with the difference between formats being statistically significant at the 99% confidence level). Even after analyzing separate groupings of semesters, the authors find that the pattern in this particular distribution of outcomes remained the same across all groupings.

In addition, although the average prior GPAs of students in the online sections were slightly higher than those of their face-to-face peers (2.81 vs. 2.76), the average GPAs earned by students in the face-to-face courses were significantly higher than those earned by students in the online sections (2.8 vs. 2.65). Nevertheless, Tanyel and Griffin find that this difference in average GPAs between delivery formats was driven primarily by courses during the last five semesters, during which students in the face-to-face sections earned an average GPA of 2.83 (compared to an average GPA of 2.66 for their counterparts in the online classes). These semesters also corresponded with enrollment more than tripling in the online sections, suggesting that there was likely more heterogeneity among the subset of students enrolled in online courses.

One of the most impressive aspects of this study is the sheer size and diversity of the number of courses surveyed that utilized at least one face-to-face and online section within a semester. It is hard to think of a study that has surveyed such a large number of side-by-side online and face-to-face courses over such a long time period (for ten years between 2002 and 2011). As a result, the authors should be commended for putting together such a rich dataset of courses that allows them to exploit variations in delivery formats and student outcomes across a very large sample of observations. Nonetheless, this study still has some methodological limitations that are worth addressing.

First, Tanyel and Griffin acknowledge that their study presents only a descriptive analysis of rich archival data and, due to their inability to control for important covariates, should not be treated as an analysis of the true “causal effects” of online learning on student outcomes. While it is understandable that the presence of sample

selection bias might exist given that students were not randomized into each section, it would have been valuable had the authors conducted a multivariate regression analysis that attempted to rigorously ascertain the effect of delivery format on student outcomes—holding everything else constant. This is particularly important, given some significant differences in average characteristics that the authors already pointed out between students in each delivery format. For example, not only do students in online courses tend to have higher average prior GPAs, but they also have higher average withdrawal rates. With regard to the latter point, an analysis of outcomes on course grades may then overestimate the true effect of the online delivery format on student outcomes. Controlling for these various effects would thereby mitigate biases like the “clientele effect” frequently referred to by the authors.

Furthermore, by presenting only aggregate statistics, Tanyel and Griffin somewhat sacrifice precision for the sake of harnessing the large sample size. While this is not unreasonable, it would have been useful had the authors presented evidence on the degree to which the structure and substance of each delivery format within a course was comparable to each other. In fact, the definition of “online” courses encompassed a wide range of technology-enhanced sections, from courses that were fully online to those that utilized more of a hybrid format. Because the authors cannot distinguish between modes of delivery, it remains difficult to determine exactly what outcomes were associated with which delivery format (and the extent to which they could derive recommendations regarding the desirability of a specific technology-infused delivery method). Finally, given that the majority of the courses surveyed were from the College of Arts and Sciences, the external validity of the results in this study might be limited only to similar settings.

Bibliography of Studies

Burns, Kathleen, Mimi Duncan, Donald C. Sweeney II, Jeremy W. North, and William A. Ellegood. "A Longitudinal Comparison of Course Delivery Modes of an Introductory Information Systems Courses and Their Impact on a Subsequent Information Systems Course." *MERLOT Journal of Online Learning and Teaching* 9, no. 4 (2013): 453-467.

Carmichael, Tami S., Jeffrey S. Carmichael, and Rebecca Leber-Gottberg. "Capstone Courses: Student Learning Outcomes Indicate Value of Online Learning in Pilot Study." *International Journal of Virtual Worlds and Human-Computer Interaction* 2 (2014): 73-82.

Fish, Kristine and Hyun Gu Kang. "Learning Outcomes in a Stress Management Course: Online versus Face-to-Face." *MERLOT Journal of Online Learning and Teaching* 10, no. 2 (2014): 179-191.

Jones, Sherry J. and Vena M. Long. "Learning Equity between Online and On-Site Mathematics Courses." *MERLOT Journal of Online Learning and Teaching* 9, no. 1 (2013): 1-12.

Jorczak, Robert, and Danielle Dupuis. "Differences in Classroom Versus Online Exam Performance Due to Asynchronous Discussion." *Journal of Asynchronous Learning Networks* 18, no. 2 (2014): 67-75.

Joyce, Theodore J., Sean Crockett, David A. Jaeger, Onur Altindag, and Stephen D. O'Connell. "Does Classroom Time Matter? A Randomized Field Experiment of Hybrid and Traditional Lecture Formats in Economics." *NBER Working Paper No. 20006* (2014).

Kwak, Do Won, Flavio M. Menezes, and Carl Sherwood. "Assessing the Impact of Blended Learning on Student Performance." *Economic Record* 91, no. 292 (2015): 91-106.

Metzgar, Matthew. "A Hybrid Approach to Teaching Managerial Economics." *e-Journal of Business Education & Scholarship of Teaching* 8, no. 2 (2014): 123-130.

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

Tanyel, Faruk and Jan Griffin. "A Ten-Year Comparison of Outcomes and Persistence Rates in Online Versus Face-to-Face Courses." *B>Quest* (2014).

Verhoeven, Penny, and Tatiana Rudchenko. "Student Performance in a Principle of Microeconomics Course under Hybrid and Face-to-Face Delivery." *American Journal of Educational Research* 1, no. 10 (2013): 413-418.

Xu, Di and Shanna Smith Jaggars. "Performance Gaps Between Online and Face-to-Face Courses: Differences Across Types of Students and Academic Subject Areas." *The Journal of Higher Education* 85, no. 5 (2014): 633-659.