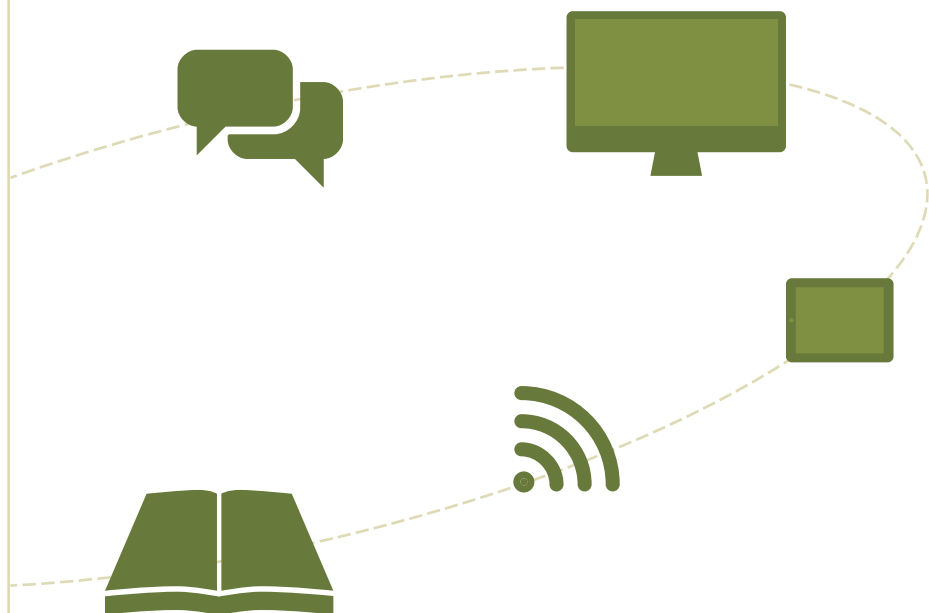




Current Status of Research on Online Learning in Postsecondary Education

Kelly A. Lack



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 2013 ITHAKA. This work is licensed under the Creative Commons Attribution No Derivative Works 3.0 United States License. To view a copy of the license, please see <http://creativecommons.org/licenses/by-nd/3.0/us>

Current Status of Research on Online Learning in Postsecondary Education

- 4 SRI/Department of Education Meta-Analysis
- 6 The NCAT Studies
- 7 Additional Studies
- 13 Conclusions and Other Observations
- 16 Studies Included in the Literature Review
- 20 Appendix
- 70 Bibliography of Studies

Current Status of Research on Online Learning in Postsecondary Education

Kelly A. Lack¹
March 21, 2013

It is unfortunately the case that there have been few rigorous efforts to produce compelling evidence of the learning outcomes associated with online courses at the postsecondary level.² While the presence of various forms of online learning has expanded rapidly over the past decade, the field has proliferated largely on the proverbial “wing and a prayer.” This is not to suggest that the future of online learning is bleak; on the contrary, the potential for online learning to prompt substantial change throughout higher education is immense. At the same time, most, if not all, of the research that has been published on this topic—which has

1 I would like to thank William G. Bowen for his assistance and support as I was preparing this literature review, as well as my colleague, Johanna Brownell, for her time and help throughout the entire process. Kevin M. Guthrie, president of ITHAKA, has also discussed the content and has reviewed the current document, as has Lawrence S. Bacow, Senior Advisor to Ithaca S+R (shorthand for “ITHAKA Strategy and Research”).

In light of the initial submission deadline for this paper, the original version of this literature review, published in May 2012, did not review studies published after the end of 2011. Ithaca S + R released an addendum to that document, in fall 2012, extending its analysis to three articles published between the end of 2011 and August 2012. The addendum also identified, but did not analyze in depth, other studies that fit the original literature review’s criteria but that were not included in the original version. This new, March 2013 version of the literature review not only incorporates the contents of the original literature review and of the addendum, but also analyzes several additional studies that meet the original literature review’s criteria and that were published no later than December 2012.

2 It is conceivable that the same conclusion holds at the K-12 level, but I have not made as comprehensive an effort at that level to chart the relevant research.

been reviewed in detail in preparing this report³—has substantial limitations and therefore is of limited value in determining how much more of an investment, and in what form, should be made in this field. As of this moment, a priori arguments will have to continue to bear much of the weight.

This report describes in detail the status of research regarding online or hybrid⁴ (part online and part face-to-face) learning at the college level, and discusses the need for more rigorous studies of the learning effectiveness and cost-effectiveness of new forms of technology-enhanced instruction.

SRI/Department of Education Meta-Analysis

Before summarizing the more recent studies, it is necessary to set a baseline by reviewing briefly the widely publicized report prepared by SRI (often cited as Means et al.), which was originally released by the U.S. Department of Education in 2009, and then revised in 2010.⁵ The first thing to be said is that this report sets a fairly high bar for including studies in its meta-analysis. The studies have to compare web-based instruction (i.e. excluding video-based courses and stand-alone computer applications) to face-to-face or “offline” instruction; they have to use random assignment or quasi-experimental research designs; and they have to focus specifically on objective measures of student learning (that is, they cannot look solely at outcomes such as student or faculty perceptions) that are measured for both the web-based and non-web-based formats. The studies also need to have been completed at the time of the meta-analysis, and have to provide sufficient information to allow effect sizes to be calculated. The meta-analysis authors begin with an initial pool of 1,132 studies published between 1996 and 2008, and they end up with just 45 that meet the specified criteria. This dramatic fall-off in numbers is itself a telling reflection of the lack of sophistication of most activity in this field. Another 77 studies without face-to-face control groups are included in a “narrative synthesis.”⁶ Of the 45 studies reviewed in detail, most have sample sizes of a few dozen learners; only five include more than 400 learners. Many of the studies are in the fields of medicine or health care; very few have direct relevance to the large public universities or the broad-based community colleges that educate such a high fraction of this country’s student population.

3 While at one point I considered including in this document an investigation of the marketing claims of providers of online learning, I ultimately concluded that any thorough effort of this kind was impractical. If literature supporting the claims of providers existed, and were deemed rigorous and reliable, it would most likely have been publicized, as doing so would be in the providers’ interests. At this time, I am unaware of any such research. An obstacle to third-party evaluation of marketing claims is the natural reluctance of for-profit providers to give independent external parties access to sensitive materials and proprietary data.

4 “Hybrid” courses are sometimes called “blended” courses. Throughout this document I will use the term “hybrid” for the sake of consistency.

5 See Barbara Means et al., *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies* (Washington D.C., U.S. Department of Education, 2010), www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf, for a revised version of this meta-analysis.

6 Means et al., p. xii.

All that said, the findings of the DOE meta-analysis are mildly encouraging for the further development of blended and purely online education.⁷ The authors find that students who took their classes in a blended format performed significantly better, on average, than did those taking courses through traditional, face-to-face instruction, and that there is no significant difference between the purely online and purely face-to-face formats. The fact that the hybrid mode is associated with significantly better outcomes is hardly surprising, since hybrid forms of instruction generally involve the commitment of more resources. Time-on-task is found to be a significant variable when studies involving K-12 education are excluded from the analysis, and effect sizes are larger when a new instructional approach is tried, as opposed to simply varying the medium of delivery. In addition, the meta-analysis finds that online learning can be enhanced by giving learners control of their interactions with delivery mechanisms and prompting learner reflection. “The meta-analysis findings do not support simply putting an existing course online,” the DOE report says, “but they do support redesigning instruction to incorporate additional learning opportunities online.”⁸

While the DOE meta-analysis has received a great deal of attention for its comprehensiveness and methodology, it has not gone uncriticized. In an NBER working paper titled “Is It Live or Is It Internet? Experimental Estimates of the Effects of Online Instruction on Student Learning,”⁹ David N. Figlio of Northwestern University and his co-authors criticize the methodology of the original (2009) DOE meta-analysis and argue that much more work—and, in particular, much more careful experimental work—needs to be done before sweeping conclusions can be reached. To demonstrate the kind of research they think is needed, Figlio and his colleagues conduct a randomized study

7 It should be mentioned that the DOE meta-analysis, while one of the most comprehensive and most well-known analyses done in the last 10 years, is not the first to document a no-significant-difference (NSD) finding for different learning formats. Also worthy of mention is a 2001 book by Thomas L. Russell, entitled *The No Significant Difference Phenomenon: A Comparative Research Annotated Bibliography on Technology for Distance Education* (2001, IDECC, fifth edition), which identifies more than 350 studies, dating as far back as research on “correspondence study” in 1928, that show no significant difference in student learning outcomes based on mode of delivery. Today, a companion website (nosignificantdifference.org) offers a searchable index of studies pertaining to the effect of delivery mode on learning outcomes, that includes research published after the release of Russell’s book. The website identifies not only studies that show no significant difference between traditional face-to-face learning and different forms of distance education (defined broadly), but also studies that show that one format is associated with better learning outcomes, or that reflect mixed results. The studies referenced on this site vary widely, in terms of the different types of delivery format involved, the age group of the participants, the methodology used, and a number of other dimensions. While the NSD website provides a useful database for research comparing different forms of instruction, it does not attempt to analyze critically the quality of any of the studies it cites; it is up to the reader to identify the literature most relevant to a particular research question and to consider the quality of each study. As the website rightly notes: “While the NSD finding may be prevalent enough across a large enough pool of MCS [media comparison studies] research results to be considered a fair conclusion despite the lack of controlled variables, other findings limited to certain studies may not be as widely applicable. Therefore, a good practice in performing meta-analyses . . . of the MCS research might be to first sort for collections of MCS that control for similar sets of variables, and then analyze the findings from those collections.”

8 Means et al., p. 51.

9 David N. Figlio, Mark Rush, and Lu Yin, “Is It Live or Is It Internet? Experimental Estimates of the Effects of Online Instruction on Student Learning,” National Bureau of Economic Research Working Paper No. 16089 (2010), <http://www.nber.org/papers/w16089> (accessed July 20, 2010).

of students in a microeconomics course and compare the learning outcomes of students who attended the live lectures and students who watched videos of the lectures online. They find no significant differences in outcomes.

Another review of the 2009 DOE meta-analysis—by Columbia University’s Community College Research Center (CCRC)—also concludes that the DOE meta-analysis contains little data relevant to the question of the effectiveness of fully online courses in typical college and university settings. Many of the studies in the DOE meta-analysis are of hybrid modes of teaching that continued to provide a substantial amount of face-to-face instruction, and many are studies of short educational programs among student populations very different from those who attend community colleges. In fact, the CCRC report says that only seven studies in the DOE meta-analysis involve undergraduate or graduate students enrolled in semester-long online courses, and these seven studies show no significant advantages or disadvantages of online courses. The authors of the CCRC report also express concern that online courses may work less well for low-income students with low-speed Internet access at home.¹⁰

A few comments on this discussion are in order. The CCRC report is correct in pointing out the limited relevance of the DOE report for typical college and university settings. Figlio and his colleagues’ own study is arguably also of limited relevance, in that it is not applicable to many varieties of online courses that involve more than just online lecture videos; however, it is still of interest in light of the proliferation and increasing popularity of video-driven courses,¹¹ and he and his colleagues deserve credit for the relative rigor of their study design. Finally, Figlio, his colleagues, and the CCRC are right in calling for much more extensive and much more careful experimental research into the effectiveness of sophisticated modes of online teaching in college and university settings.

The NCAT Studies

A further criticism of the existing research pertaining to postsecondary education is that, with a few notable exceptions, little attention has been paid to the costs associated with different delivery formats—and those cost studies that do exist have severe limitations. In this regard, there is another large set of data regarding online learning initiatives that, despite its defects, deserves

10 Shanna Smith Jaggars and Thomas Bailey, *Effectiveness of Fully Online Courses for College Students: Response to a Department of Education Meta-Analysis*, Community College Research Center, Teachers College, Columbia University (July 2010), <http://ccrc.tc.columbia.edu/Publication.asp?UID=796> (accessed July 20, 2010).

11 Perhaps the most prominent current examples of online instruction involving online videos are Khan Academy and massive open online courses (MOOCs), though in many cases these online materials are provided outside of a traditional university setting. Khan Academy is a nonprofit organization that is perhaps best known for its short instructional videos, hosted on YouTube, on topics ranging from civics and art history to computer science, chemistry, differential equations, and the Greek debt crisis. Students who watch Khan Academy videos generally do not receive university credit. MOOCs are usually taught to thousands of students worldwide by well-known professors. Students who register for MOOCs (typically for little or no charge) watch videos and complete assignments that are machine-graded or graded by other students or teaching assistants. MOOCs also sometimes offer online discussion boards or other opportunities for students to interact with each other in a virtual setting. Some MOOCs lead to certificates of accomplishment or badges (which sometimes require payments of a modest fee); few MOOCs carry college credit.

careful attention. Carol Twigg and her colleagues at the National Center for Academic Transformation (NCAT) have worked with well over 70 campuses to introduce course redesigns that involve a variety of models of online instruction. The NCAT reports regarding these initiatives describe each one in considerable detail and (unlike most other studies) document cost savings in essentially every case; they also generally report some improved student learning outcomes.¹²

However, these case studies have two serious limitations from a research standpoint. First, they are all “before-and-after” studies, which compare the results achieved by a redesigned course in one or more years with the results for the traditional variant of the same course in a previous year (or years). Controls are lacking, and one of NCAT’s quarterly newsletters is frank in acknowledging that in some cases situations have changed noticeably between the “before” and “after” years in question.¹³ These are not the kind of rigorous, randomized studies that can produce robust results. Claims about the success of NCAT transformations are typically based on case studies and examples, and the results are all self-reported; careful third-party evaluation would be preferred.

Still, the consistency of the reported outcomes certainly strongly suggests that redesigning courses along the lines suggested by the various NCAT models can produce positive results. Twigg deserves great credit for having seen the potential of online and hybrid instruction early on, having had the courage to emphasize the importance of achieving cost savings, and having worked hard (although, arguably, without marked success) to encourage both replication and sustainability.¹⁴

Additional Studies

As the popularity of web-based learning continues to escalate—as a result of a combination of factors including, but not limited to, increasing familiarity with forms of technology among college-age students, and budgetary pressures to teach more students with fewer resources—examining what others have learned about the effectiveness of web-based courses is helpful, even if great caution is warranted in interpreting some, if not all, of their findings.

Although the 2009 DOE meta-analysis analyzes a substantial body of research on the relative effectiveness of web-based instruction and traditional, face-to-face learning, by no means can its review of the literature on this subject be considered

-
- 12 An Education Sector report (Ben Miller, *The Course of Innovation: Using Technology to Transform Higher Education*, May 2010) provides a very useful summary of NCAT’s work, as well as references to all of the underlying NCAT reports conducted up to the time of its preparation.
 - 13 See NCAT, “The Learning MarketSpace,” July 2010, <http://www.thencat.org/Newsletters/Jul10.html>, especially p. 3.
 - 14 Both Miller’s report and a subsequent column by Kevin Carey (“The Better Mousetrap Problem,” *Chronicle of Higher Education*, (May 27, 2010)) point out that, in Carey’s words, “in many ways the most interesting part of the NCAT story isn’t the colleges that have adopted these proven methods. It’s the colleges that haven’t—i.e., most colleges” (out of a universe of many thousands nationwide). Both Miller and Carey discuss the reasons for this pattern in detail. In addition to the lack of evidence that the courses redesigned at one institution have been adopted elsewhere—the math “emporium” work done at Virginia Tech being one exception—it is unclear how many of the redesigned courses have been sustained at their home campuses. Some case studies done by my ITHAKA colleague, Matthew P. Long, suggest that it has been difficult in some settings to upgrade technology and to sustain the original efforts.

exhaustive. In the years following the cut-off date for inclusion in the DOE paper, several other studies (including the one by Figlio and his colleagues, cited earlier) have been published that compare web-based and face-to-face learning in undergraduate courses. A search for other literature on this topic yielded still more studies that may have been available to the authors of the DOE meta-analysis at the time the meta-analysis was written but were not included for some other reason (for example, perhaps they were simply overlooked).

This section of the report considers studies regarding web-based and face-to-face learning that were not included in the DOE meta-analysis and that satisfy a set of basic criteria. To be included in this follow-on literature review, a study must fulfill all of these requirements:

1. Compare face-to-face learning to hybrid or online learning.
2. Examine learning outcomes or academic performance (as opposed to comparing delivery formats based only on self-reported outcomes such as perceived learning or student satisfaction with the course).
3. Involve one or more undergraduate, for-credit college course(s) that are offered outside of a continuing education program.
4. Take place in the United States or in a location with a comparable culture and higher education system (such as England or Australia).
5. Be authored by someone who is not a current student (for instance, a paper submitted in partial fulfillment of dissertation or bachelor's degree requirements would be excluded from the review).

Approximately 30 studies were identified that fulfill these criteria. (Please see the Table at the end of this paper, which contains information on each study, as well as the Appendix, where each study is described in even more detail.) The subject matter of the courses involved ranges from marketing and accounting to psychology and English composition, from statistics and probability to biology and electrical engineering. Some studies were conducted at community and technical colleges; others took place at private and public four-year colleges and universities in various regions of the country.

Cautions

A few words of caution are necessary. Very few of the studies examined involve a rigorous design in which students were randomly assigned to the face-to-face format or the online/hybrid format¹⁵; instead, students were generally allowed to self-select into a course format. Furthermore, of the observational studies, some do not make any attempt to control for pre-existing differences in the students taking the online or hybrid version course and the students taking the face-to-face version. This limitation is particularly noteworthy because, in a number of studies, the differences in background characteristics between the two groups are not trivial. Moreover, of the few studies that do include a rigorous

15 As noted in the following paragraph, some studies in the literature review compare the face-to-face version of one or more courses to a purely online version of the same course(s); others compare the face-to-face format to the hybrid format; and still others compare the face-to-face format to both the hybrid format and to the completely online format. Unfortunately, there are also some studies where the comparison is unclear, as the authors fail to specify the degree of face-to-face contact students in the web-based format had with their instructor.

Very few of the studies examined involve a rigorous design in which students were randomly assigned to the face-to-face format or the online/hybrid format; rather, students were generally allowed to self-select into a certain course format.

design, most do not have a large sample size; of the three studies¹⁶ that employ randomization (the “gold standard,” so to speak, in evaluation research), the largest sample size consists of only 36 students across the two formats.¹⁷

In addition to the variety in subject matter of the courses involved in the studies, the studies differ in whether the authors are comparing traditional courses (with one or more face-to-face meetings each week) with purely online courses (with no face-to-face meetings), or whether they are comparing traditional courses with hybrid courses, which involve a combination of web-based learning and face-to-face meetings. However, what is meant by an “online,” “web-based,” or “hybrid” course in one study is not necessarily the same as what is meant by an “online,” “web-based,” or “hybrid” course in another. Some of the online courses in the studies provided students with instant feedback and hints as they worked through activities or simulations, which were supplemented by virtual interaction, via discussion boards and chat rooms, with other students and the instructors. Other courses seem far less sophisticated and appear to consist of little more than video recordings of lectures posted on the Internet. The variety of courses available that can be described as “online” or “hybrid” is noteworthy because the differences in their features can play a significant role in how much students learn from those courses. The same group of students may do much better than students in the face-to-face comparison group when the former group is using a highly sophisticated online approach, but much worse than the comparison group when it is are using a relatively rudimentary online approach. While some studies give an in-depth description of what the “online” course being evaluated entailed (in some cases, even providing screen shots or sample exercises), others do little to elaborate on what is meant by “online” or “hybrid.”

Some studies also involve a comparison of multiple face-to-face courses with the online or hybrid versions of those same courses, using final course grade or completion rate as the only performance-related outcomes. In such studies it is not always clear how much the “online” courses differ from each other. If the online courses in the study are relatively similar to each other—for example, if the same learning management systems were used, the same types of homework assignments were given, and the same types of automatic feedback were provided

16 The three randomized studies to which I refer are those conducted by Poirier and Feldman (2004), Mentzer, Cryan, and Teclehaimanot (2007), and Olson (2002). While a fourth study, by Christou, Dinov, and Sanchez (2007), is described in its abstract as a “randomized controlled study,” there is no mention, in the section of the article where the authors describe the study’s design, of randomly assigning students, rendering it unclear whether random assignment of students actually took place.

17 My colleagues and I conducted a randomized study between fall 2010 and spring 2012 to help fill this void in large-scale rigorous research about web-based learning in higher education. In our study, we randomly assigned more than 600 students on six public college campuses to take either a hybrid or purely face-to-face version of an introductory statistics course. At the end of the semester, we used regression analysis to compare the performance of the students in the two formats with respect to pass and completion rates, standardized test scores, and scores on a common set of questions on students’ final exams, incorporating into our analysis background data collected from the campuses’ institutional research offices. As my colleagues and I learned, however, implementing a randomized study is both expensive and extremely difficult to implement on college campuses, in part because of Institutional Review Board requirements, logistical challenges related to students’ registering for the appropriate format and section of the course, some students’ reluctance to comply with their random assignments, and a variety of other issues. (See William G. Bowen, Matthew M. Chingos, Kelly A. Lack, and Thomas I. Nygren, *Interactive Learning Online at Public Universities: Evidence from Randomized Trials*, May 22, 2012, available on the ITHAKA website at www.sr.ithaka.org.)

to students by the machine—the results might indicate how effective that particular type or model of online learning is, relative to the face-to-face format. On the other hand, there may be much more heterogeneity in the types of online or hybrid courses being studied: for example, the online version of introductory statistics may be much different from the online version of introductory economics (in other ways than just subject matter). If there is substantial heterogeneity, the studies may not be measuring the effectiveness of a particular kind of online learning so much as evaluating the effectiveness of the “average” (so to speak) online course included in the sample. It is useful to know how much and what kind of variation exists among the different forms of web-based instruction involved in the study, but unfortunately this information is not always available.

The same can be said about the face-to-face instruction involved in the studies in this review. Large (or small) variations across face-to-face sections may exist in terms of the amount of weekly face-to-face interaction students had with instructors, class size, experience or skill level of the instructor, how any face-to-face time was used, and so on. Unfortunately, information about what face-to-face instruction in the relevant sections involved, and how much variability there was across those face-to-face sections, was not provided in some studies (much less incorporated into the authors’ analyses).

The limitations described above are indicative of general patterns in research about “online learning.” Each study included in this literature review also has more specific limitations, suggesting that, when considered alone, that particular study cannot provide conclusive evidence about the effectiveness (or ineffectiveness) of online (or hybrid) learning.¹⁸

Findings

A holistic look at the literature assembled, bearing in mind the caveats described above, yields little, if any, evidence to suggest that online or hybrid learning, on average, is more or less effective than face-to-face learning. Not only do the types of online or hybrid learning involved in the studies in this literature review vary considerably, but so do the kinds of outcomes measured, which range from homework assignment scores and project grades, to exam scores, final course grades, and completion and withdrawal rates. Most studies employ multiple measures of student performance.

Even assuming that the measures employed are useful indicators of the student learning that takes place in the respective courses, the literature does not show a pattern of online students consistently outperforming their face-to-face-format peers, or vice versa. Even within a single study, there are few instances in which one group outperforms the other group on all performance measures evaluated

¹⁸ See the Appendix for a more detailed description of each study.

(Xu and Jaggars' longitudinal study being one of the few exceptions¹⁹). Most of the studies have mixed results: on some of the measures, the students in the online or hybrid format did better, but on others they did worse, relative to students in the face-to-face format—or else, on some measures the online- or hybrid-format students did significantly better or worse than the students in the face-to-face format, but on other measures there was no significant difference between the two groups.²⁰

The following summary illustrates the range of results from the studies in this literature review and the variety of methodologies used in those studies:

- Of the three studies in which the researchers report randomly assigning students to a face-to-face or online format, one study shows no significant difference in exam performance between online and face-to-face students, but online students were less likely to submit assignments and consequently earned significantly lower course grades (Mentzer, Cryan & Teclhaimanot, 2007). The second study shows that online students performed significantly better on exams, but about the same on papers (Poirier & Feldman, 2004). The third study shows no significant difference between the two formats (Olson, 2002). However, while these three studies have the most rigorous designs of the studies examined in this review, they unfortunately also have very small sample sizes: each study involved fewer than 40 participants.
- Of the studies that include more than 1,000 participants taking a variety of courses that were offered both in online/hybrid (the studies do not always distinguish between the two) and face-to-face formats, one study (Xu & Jaggars, 2011) shows that students taking online/hybrid courses had a higher probability of withdrawing and received lower course grades than students in face-to-face courses. By contrast, Carpinelli et al. (2006)

19 In February 2013, Xu and Jaggars of the Community College Research Center at Columbia University's Teachers College released another study that involved a similar but larger cohort of community college students. Like the 2011 longitudinal study cited above, the 2013 study involves students who enrolled in community and technical college students in Washington State in fall 2004, and who were tracked through spring or summer 2009; the new study analyzes data from more than 40,000 community college students who collectively took about 500,000 courses. Unlike the 2011 study, however, the 2013 study is focused on comparing performance in online and face-to-face courses by the same student (to measure what the researchers called the student's "adaptability to online learning"), and it examines grades from all courses taken by all degree-seeking students. (By contrast, the 2011 study focuses only on transfer-oriented students and excludes remedial courses from the analysis.) Like the 2011 study, the results of the 2013 study show that students generally earned lower grades in online courses than in face-to-face courses; it also shows, though, that some groups of students (for example, older students and female students) "adapted" better than others (for example, younger students and male students). However, given that, before this February 2013 CCRC study was released, I had already decided to limit my review of the literature to studies published through the end of 2012, I do not describe this new CCRC study in detail. (See Di Xu and Shanna Smith Jaggars, "Adaptability to Online Learning: Differences Across Types of Students and Academic Subject Areas," CCRC Working Paper No. 54, Community College Research Center, Teachers College, Columbia University (February 2013), <http://ccrc.tc.columbia.edu/publications/adaptability-to-online-learning.html> (accessed February 20, 2013).)

20 Throughout this document, I use 10 percent as a threshold significance level; that is, when I say a statistic is not significant, I mean that the p -value is greater than 10 percent. Not all of the authors of the studies reviewed in this document, however, use a 10 percent threshold; some authors use a 5 percent threshold, thus only considering a result significant if the p -value is 0.05 or smaller. Whenever possible, I will specify the p -value of any finding described, so that the reader may make his own determination about whether a given result should be considered significant.

find that “distance learning” students earned significantly higher course grades. A third study shows that students had higher grades in the face-to-face sections of some community college science courses but not in other community college science courses, and that when a subset of the community college participants transferred to four-year public institutions, there was no significant difference in the students’ grades in science courses taken at those four-year institutions (Colorado Department of Higher Education, 2012).

- Of the studies that do not involve randomization (and thus whose results are subject to self-selection bias), but whose analyses include controls for a number of background characteristics and/or other predictor variables, some offer results that are mildly encouraging, though not conclusive, for hybrid or online formats. Christou, Dinov, and Sanchez (2007), for example, find that the overall performance of hybrid students was significantly better than the overall performance of face-to-face students. Other studies have mixed results, or results that show no significant difference between the two formats. In Bennett et al.’s (2007) research, students in the face-to-face format achieved significantly higher averages than their counterparts in the online format in one economics course, but the reverse was true in another economics course. In addition, the regressions by Ary and Brune (2011) that control for background characteristics show that the percentage change in score between pre- and post-tests was significantly higher for face-to-face students, but that there was no significant difference between online and face-to-face students with respect to course averages. When they control for background characteristics, Weber and Lennon (2007), Daymount and Blau (2008), Wilson and Allen (2011), and Driscoll et al. (2012) each find no significant difference between online and face-to-face students. Still other studies show no significant difference in the aggregate, but that particular populations performed significantly better or worse in one format or the other. For example, Riffell and Sibley (2005) find that hybrid-format students did better than face-to-face students on post-course assessments, but that the difference was only significant for upperclassmen. Research by Friday et al. (2006) shows no significant difference in grades in two courses when the results are aggregated, but that, when the data are analyzed by gender, men did worse in the online format in one of the two courses.
- Other studies are strictly observational, with relatively small sample sizes and few (if any) attempts to control for student background characteristics. Some of these studies involve some degree of standardization of instruction between the web-based and face-to-face formats; others do not. Of the research that falls into the former category, the findings tend to indicate no significant difference, or provide mixed results. For example, Summers, Waigandt, and Whittaker (2005), Lewis and Harrison (2012) and Enriquez (2010) each find no significant difference with respect to any of the outcomes they examine. Plumb and LaMeres (2011), Rich and Dereshiwsky (2011) and Parsons-Pollard et al. (2008) each find online students outperformed face-to-face students on some measures, face-to-face students outperformed online students on other measures, and there is no significant difference on other measures. As is true with respect to the studies described in the previous paragraph, some results show differences only for certain subgroups

of students. For example, Urtel (2008) finds no significant difference between online and face-to-face instruction for male students or for black or Hispanic students, but white students and female students did better in the face-to-face format; by contrast, Wagner, Garippo, and Lovaas (2011) find no significant difference in final grades when looking at all students, but males as a subgroup did significantly worse in online classes. Finally, the studies that are observational without any attempts to standardize course materials, assignments, assessments, or other components of instruction show either no significant difference in outcomes (Signor, 2003), or else significant differences for some courses but not for others (Samman et al., 2012).

The description above is intended to provide an (incomplete) overview of the studies reviewed; please see the Table and the Appendix for more details.

By and large, these studies offer little evidence to support conclusively the claim that online or hybrid learning is significantly more effective or significantly less effective than face-to-face instruction. As is so often true, one must be wary of assertions of the unmitigated success or failure regarding online learning. It is also important, particularly in examining research pertaining to online learning, to bear in mind the variety that exists both in web-based instruction and in studies attempting to evaluate it.

Conclusions and Other Observations

The New York Times' dissection of the Carnegie Learning K-12 math course in fall 2011, as well as news coverage regarding very low graduation rates at some for-profit universities, raise questions of the possibility that providers of college-level online courses have exaggerated their success.²¹ This danger always exists, perhaps especially where for-profit providers are (understandably) seeking revenues and returns on their investments, and skepticism is warranted concerning any analysis of learning outcomes that has been done by self-interested stakeholders. It is asking a great deal of those who are basing their careers or businesses on successful adoption of online education to be objective in evaluating the success of their own products; it is relevant to note that drug companies contract out clinical trials. It is difficult to understate the value of rigorous evidence from third parties on learning outcomes, or the need to avoid assuming that what sounds exciting and promising is necessarily proven.

Yet it is also important to understand how challenging it is to do rigorous research on educational outcomes, especially where human subjects and Institutional Review Board requirements are involved. Barriers to randomization, including resistance from faculty, students, and administrators, can be difficult to overcome, and implementing proper research protocols is not easy. Still, there is no substitute for rigorous research, and Ithaca S+R has engaged in an effort—which, while it is by no means flawless, represents an

21 Trip Gabriel and Matt Richtel, "Inflating the Software Report Card," *New York Times*, October 8, 2011; Tamar Lewin, "Report Finds Low Graduation Rates at For-Profit Colleges," *New York Times*, November 23, 2010.

improvement over much of the existing research—to fill at least a small part of this void.²² Absent such research, there is no way to answer definitively the question about the prevalence of “inflated report cards.”

In addition to the need for reliable research on the effectiveness of different instructional modes, one cannot disregard the reality that, especially at the postsecondary level, individual institutions have (or should have) both a strong interest in measuring costs and a strong incentive to control or reduce costs. Thus, in assessing learning effectiveness at the college level, close attention should be paid to outcomes in relation to cost—achieving the same outcomes at half the cost per student should be seen as a nontrivial accomplishment. Carol Twigg of NCAT understands this, but many creators and users of online courses do not. A huge complication is that it is by no means easy to study costs, because there are potentially large differences between start-up or one-time costs of sophisticated online instruction, and costs in some kind of equilibrium state years down the road. As part of the companion study referenced earlier, Ithaka S+R has explored a cost simulation approach to analyzing potential cost savings, but much more work needs to be done in this area.²³

To return more directly to the question of how to respond to arguments over “inflation of report cards,” one clear lesson is to avoid claiming too much, unless or until the evidence is clear. Not going out on the proverbial limb is the best way to reduce the risk that one will fall when the limb is cut off. The ambiguity of evidence about learning outcomes may actually seem less surprising when one considers how little is known, more generally, about how college students learn.

Finally, it is important to remember that the lack to date of compelling quantitative evidence of learning effectiveness (with proper controls in place) does not imply that the rapidly increasing number of efforts to employ online learning is a mistake—not at all. The consistency of findings in even imperfect studies of learning outcomes—arguably supplemented by a priori reasoning and anecdotal commentaries—suggests that determined efforts should be made to improve online learning, taking full advantage of features that are specific to an online environment, such as automatic and interactive feedback loops. Well-designed online instruction has great promise.²⁴ That being said, there is an undeniable need for rigorous efforts to assess readily measurable outcomes such as completion rates and time-to-degree, paying attention to differences in outcomes among different student populations and subgroups. In addition, the higher education community would benefit from careful analysis of the costs associated with technology-enhanced forms of instruction, especially in

22 See Bowen et al., May 2012, *Interactive Learning Online at Public Universities*—with particular attention to Appendix C—for a discussion of the difficulties of conducting this kind of research.

23 See Bowen et al., *Interactive Learning Online at Public Universities*.

24 VanLehn’s meta-analysis of randomized studies concerning the effectiveness of synchronous, computer-based tutoring may also be of interest to the reader in this regard. While not directly within the scope of this literature review, VanLehn’s analysis suggests that, in the fields of science, technology, engineering, and math, the instructional approaches used by some “intelligent tutoring systems” can be as effective in some settings as one-on-one tutoring by a human. See Kurt VanLehn, “The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems,” *Educational Psychologist*, 46.4 (2011): 197-221, <http://dx.doi.org/10.1080/00461520.2011.611369>.

a long-term equilibrium state. Potential cost savings are far from inconsequential, especially in sectors of postsecondary education where there is unrelenting pressure to do more with less.

STUDIES INCLUDED IN THE LITERATURE REVIEW

Study	Sample size	Institution	Course Subject	Design	Performance Outcomes	Results
Ary & Brune (2011)	91 online students, 94 FTF students	Small private university in Arkansas	Personal finance course	Quasi-experimental	Percentage change in score between pre- and post-tests and course semester average	<p>Without controls for background variables, pre-post test pct. improvement was higher for FTF students for 2 of 3 semesters and overall; in the third semester online students had greater improvement (sig. not specified).</p> <p>When regression analysis with controls for various background variables was used, pre-post test improvement for FTF students was sig. higher.</p> <p>Students in online classes had slightly higher course avgs. in 2 of 3 semesters, as well as overall; in third semester FTF students had greater avgs. (sig. not specified).</p> <p>When OLS regression analysis was used to compare course avgs. no sig. diff. between formats.</p> <p>Males in the FTF format had higher semester avgs. overall and greater pre-post-test pct. change overall; variation across semesters (sig. not specified).</p> <p>Females in the FTF format had greater pre-post-test pct. change overall, but lower semester avgs. overall; variation across semesters (sig. not specified).</p>
Bennett et al. (2007)	92 online students, 406 FTF students	Jacksonville State University	Principles of macro- and microeconomics courses	Quasi-experimental	Final course grade for students completing courses	<p>When looking at all econ. students, no sig. diff. in final avg. between online and FTF students.</p> <p>When students separated into micro and macro classes, FTF micro students had significantly higher final avgs. than online micro students, but online macro students had sig. higher final avgs. than FTF macro students.</p>
Carpinelli et al. (2006)	7,701 course enrollments (1,683 DL enrollments, 6,018 FTF enrollments)	New Jersey Institute of Technology	150 courses	Observational	Final course grade	Students in distance learning courses earned sig. higher grades than students in FTF courses.
Christou, Dinov & Sanchez (2007)	Stats 13 study: ≈160 total students; Stats 100A study: 39 hybrid students, 34 FTF students	University of California, Los Angeles	Introductory probability (Stats 100A) and introductory statistics (Stats 13)	Experimental (according to abstract) with standardization in instruction ¹	Scores on exams and overall performance	<p>For Stats 13: Overall performance sig. greater for hybrid group. Hybrid group did sig. better on 2 of 5 tests and significantly worse on 1 of 5.</p> <p>For Stats 100A: Hybrid group scored sig. higher on 3 exams and in overall performance.</p>
Colorado Department of Higher Education (2012)	Comparing performance in community college courses: 2,395 online students, 2,190 FTF students; comparing performance in 4-year college courses: 107 online students, 337 FTF students	Colorado Community College System and 4 four-year Colorado public institutions	6 100-level courses in biology, chemistry, and physics classes	Observational	Final course grade	<p>With respect to grades earned in community college science courses, FTF students had sig. higher grades.</p> <p>Looking at data by subject, FTF students had sig. higher grades in chemistry and biology courses; no sig. diff. in physics courses.</p> <p>With respect to performance of a subset of participants at 4-year institutions, no sig. diff. by format, either in aggregate or in individual subjects.</p>
Daymout & Blau (2008)	64 online students, 181 FTF students	Large public university in large eastern metropolitan area	Organization and Management	Quasi-experimental	Average quiz score and final course grade	<p>When no controls for background variables included, no sig. diff. in final course grades, but online group had sig. higher avg. quiz score.</p> <p>When controls added, no sig. diff. in either avg. quiz score or final course grade.</p>

Dell, Low & Wilker (2010)	13 online students, 46 FTF students	[not specified]	Educational psychology	Observational with standardization in instruction	Scores on 2-3 page report on experiment, on analysis of learning episode, and on essay-format question on final exam	Online students did sig. better on report on experiment; no sig. diff. between two formats in scores on analysis of learning episode or essay-format question.
Dinov & Sanchez (2006)	20 hybrid students, 39 FTF students	University of California, Los Angeles	Introductory probability	Observational with standardization in instruction	Scores on HW, midterm exam, and final exam; course grade	UGs in hybrid group did slightly better on all measures than UGs students in FTF group, but results reversed when graduate students included in analysis.
Driscoll et al. (2012)	170 online students, 198 FTF students	"Large urban university" in North Carolina	Introductory course cross-listed in sociology and anthropology departments	Quasi-experimental with standardization in instruction	Test score and score on an "integrating data analysis" (IDA) assignment	With no background variable controls, students in the online format performed sig. worse on test than FTF students, but when GPA was controlled for, impact of format on test performance disappeared. IDA assignment: when no covariates were included, online students did sig. worse, but when student GPA was included as covariate, effect was no longer sig.
Enriquez (2010)	25 online students, 30 FTF students	Cañada College (member of California Community Colleges)	Circuits analysis	Observational with standardization in instruction	Scores on 15 HW assignments, 4 tests, and final exam; retention rate; proportion of students earning A, B, or C in course	Retention rate and pct. of students who received an A, B, or C was the same across two groups. Online students did slightly better with respect to HW assignments, final exam, and course grade, and slightly worse on tests, but diff. are not sig.
Friday et al. (2006)	678 online students, 669 FTF students	Commuter university in major metropolitan area in Southeastern U.S.	Organization and Management and Strategic Management	Quasi-experimental with standardization in instruction	Final course grade	No sig. diff. in final course grade between two formats for either management course. In Strategic Management, men in online sections received lower grades than women in online sections and lower grades than either gender in FTF sections. (No format-gender interactions in Organization and Management course.)
Holdhusen (2009)	10 online students, 20 synchronous audiographics students, 23 hybrid students, 14 FTF students	Unspecified, though author affiliated with University of Wisconsin	Engineering graphics	Observational with standardization in instruction	Scores on post-test and on very similar lab exercises	Synchronous audiographics section did the best overall on the post-test; online section did the worst. The only sig. diff. between 2 formats was between hybrid and audiographics (the latter did better). Score of FTF students on post-test not sig. diff. from score for any other delivery mode.
Lewis & Harrison (2012)	32 online students, 28 FTF students	Small liberal arts college	Introductory social science course	Observational with standardization in instruction	Scores on 2 in-class quizzes, on 3 in-class exams, and on final exam with post-test embedded	No significant difference on post-test or 2 exams. Hybrid students did better overall with respect to final exam, total course grade (which excluded pre- and post-test scores), both quizzes, and on 1 other exam.
Mentzer et al. (2007)	18 online students, 18 FTF students	Medium-sized campus, with at least 12,000 students	Early childhood education survey course	Experimental with standardization in instruction	Scores on midterm and final exams and final course grade	Equivalent performance on midterm and final exams, but final grades were sig. lower for online students, who were less likely to complete other assignments that counted toward final grade.
Olson (2002)	17 online students, 18 FTF students	Western Oregon University	Introduction to Computer Science	Experimental	Completion rate and measure of student "success" based on multiple-choice quiz scores	No sig. diff. in completion rates.
Parsons-Pollard, Lacks & Grant (2008)	425 online students, 305 FTF students	Large, urban public university in southeast with total enrollment of 30,000	Introductory criminal justice	Observational with standardization in instruction	Final course grade (based on 3 exam grades)	Only performance on 3rd exam and final course grade were sig. correlated with format. Online group did better on exam, but FTF group had sig. higher final grades.

Plumb & LaMeres (2011)	≈ 30-40 students in each format (online and FTF)	Montana State University	Introduction to Logic Circuits	Observational with standardization in instruction	Scores on 8 quizzes; on HW assignments; on exams; course grade	No sig. diff. in course grades. Sig. diff. in scores for 2 quizzes: on one, online group did better, but on other, FTF group did better.
Poirier & Feldman (2004)	12 online students, 9 FTF students	Large state university	Introductory psychology	Experimental with standardization in instruction	Grades on 4 exams and 2 papers	Online students did sig. better on final exam; no sig. diff. in paper grades between two formats.
Reasons, Valdares & Slavkin (2005)	76 online students, 119 hybrid students, 208 FTF students	University of Southern Indiana	Introduction to Educational Psychology and Introduction to the Health Care Delivery System	Observational with standardization in instruction ²	Final course grade and quality of course participation	Online students earned sig. higher course grades than both FTF and hybrid students but no sig. diff. between grades of FTF and hybrid students. With respect to the “strength” of students’ participation levels, no sig. diff. based on the format of the course.
Rich & Dereshiwsky (2011)	19 online students, 52 daytime FTF students, 30 evening FTF students	Central Connecticut State University	Intermediate accounting	Observational with standardization in instruction	Scores on 9 online HW assignments and on short essay	No sig. diff. in HW scores, with one exception: on one topic covered in course, evening FTF students in the evening did better than daytime FTF or online students. Online students did as well as daytime FTF students on all topics. Online students also performed better overall than daytime and evening FTF students on essay.
Riffell & Sibley (2005)	55 hybrid students, 74 FTF students	Michigan State University	Introductory environmental biology	Quasi-experimental with standardization in instruction	Scores on active-learning activity and on post-test	Hybrid group did better on post-test; diff. was sig. for upperclassmen but not for freshmen; hybrid students did better than FTF students on post-test questions on topics covered by active-learning exercises. No covariate-delivery format interactions were sig.
Samman et al. (2012)	698 online students, 3,139 FTF students	Four-year public institution	Introduction to Biology, English Composition, and Fundamentals of Public Speaking	Observational	Pass rate and final course grade	With respect to final course grades, no sig. diff. between two formats for public speaking course; for composition and biology courses, FTF students did significantly better than online students. Overall FTF students had higher pass rates, but there was some variation across courses. In biology, online pass rate increased between 2008 and 2010, starting lower than FTF pass rate but ultimately surpassing it. Public speaking course pass rate was slightly lower in online course between 2008 and 2010. In composition course, the online pass rate increased between 2008 and 2010 but was always lower than FTF pass rate.
Schwartz (2012)	372 tests taken by online students, 189 tests taken by FTF students ³	National University in San Diego	Various accounting courses included in B.S. in Accountancy program	Observational with standardization in instruction	Standardized test score for many, but not all, the of the courses involved in the study, and course grade	Mean course grades and mean standardized test scores overall were sig. lower for online students. When data were disaggregated by subject area, no sig. diff. for courses in 3 of 4 areas. The exception was income-tax courses, where online students did worse with respect to both tests and course grades; when income-tax courses were excluded, no sig. diff. in two formats for test scores or course grades.
Signor (2003)	[not specified]	Swinburne University of Technology	Database Concepts and Modelling	Observational	Raw score on final exam	No sig. diff. between two formats.
Summers, Waigandt, & Whittaker (2005)	17 online students, 21 FTF students	School of Nursing at a large Midwestern university	Introductory statistics course	Observational with standardization in instruction	Cumulative score based on 3 “class exams” and on score on final exam	No sig. diff. between two formats.
Urtel (2008)	269 online students, 116 FTF students	Mid-western urban university	[not specified]	Observational with standardization in instruction	Final course grades and DFW rate ⁴	Among white students, FTF students did significantly better than online students. No. sig. diff. by format for black or Hispanic students. Female FTF students did sig. better than female online students. No sig. diff. between two formats for males.

Wagner, Garippo & Lovaas (2011)	171 online students, 435 FTF students	[not specified, though all authors affiliated with Niagara University]	Introductory course in use of business application software	Observational with standardization in instruction	Final course grade	No sig. diff. in final grades when looking at all students, but males as a subgroup did sig. worse in online class.
Weber & Lennon (2007)	17 online students, 31 FTF students (Study 1); 34 online students, 35 FTF students (Study 2)	Two private southeastern universities	Principles of Marketing	Quasi-experimental with standardization in instruction	Drop-out rate, scores on final exam and on semester project, and overall course grade	Online group had higher drop-out rate and lower project grade, final exam grade, and course grade in both Study 1 and Study 2, but when regressions were run to control for predictor variables, course format was not a sig. predictor of "learning achievement."
Wilson & Allen (2011)	58 online students, 43 FTF students	Historically black college	Principles of Management and Organizational Behavior	Quasi-experimental	Final course grade for students who completed courses	No sig. diff. in course grades between two formats.
Xu & Jaggars (2011)	18,896 students total; 126,317 course enrollments total	Washington Community and Technical College system	Multiple non-remedial courses	Quasi-experimental	Withdrawal rate and course grade	OLS assessments show online students more likely to withdraw; among students who finished course, online grades lower (controlling for student characteristics and including fixed effects). IV approach shows slightly larger diff. (i.e. OLS estimates might understate diff.). Males, younger students, and students with lower GPAs faced sig. greater penalty for taking online courses.

1 While the abstract describes the study as a "randomized controlled study," there is no mention of randomly assigning students to one format or the other in the text of the document.

2 However, while exams and assignments were standardized across the hybrid and online sections, there was no indication that this standardization occurred across the hybrid and face-to-face sections, or across the online and face-to-face sections. Thus, the standardization that occurred was of limited use for the purposes of this literature review.

3 These figures—which reflect the number of standardized tests administered in each format—are not the same as the numbers of study participants in each format. Schwartz explains: "Since many of the students who took the Intermediate Accounting tests were the same students who took tests for the Tax, Cost/Managerial, and Auditing courses, the numbers 189 and 372 represent the number of tests administered, rather than the number of different students who took the tests. Since two to three onsite sequences and two online sequences of B.S. in Accountancy courses are offered each year, it is estimated that over the two-year period for which SLOAT [standardized test] scores were analyzed, the number of different students who took these tests was approximately one-quarter the number of tests administered."

4 Rate of students earning a D, an F or a W—for withdraw—from a class.

Abbreviations

FTF = face-to-face

diff. = difference(s) or different

OLS = ordinary least squares

DL = distance learning

HW = homework

IV = instrumental variable

sig. = significant or significance

UG = undergraduate

pct. = percentage

GPA = grade point average

Appendix



Ary & Brune (2011)

Over the course of three semesters, Ary and Brune compare the performance of students in online and face-to-face personal finance classes at a small private university in Arkansas. In the three online sections involved in the study, 91 students (with 26 to 32 students in each section) completed all their coursework online, with the exception of a pre-test and post-test, a midterm exam, and a final exam (plus an optional face-to-face review session before each of these exams), all of which were given on campus. For the students in the online group, the final and midterm exams were each worth 200 points and together constituted about 50 percent of the course grade. By contrast, the 94 participants in the three face-to-face sections involved in the study (each of which had 26 to 32 students) completed five exams over the course of the semester, including a final exam; each exam was worth 100 to 110 points. For students in the face-to-face sections, exam performance constituted between 54 percent and 58 percent of the course grade.

Students in the online and face-to-face sections took the same pre-test and post-test; both groups had a personal finance simulation, online open-book quizzes, and a “track-your-spending-for-a-month” assignment. The instructors placed little weight on assignments that were not the same across groups (these assignments only constituted six to seven percent of the total points possible for the final course grade).

The results of this study reveal that, when Ary and Brune do not control for any background variables, the percentage improvement between the pre- and post-tests is higher for students in the face-to-face group for two of the three semesters, as well as for the three semesters combined. (Students in the face-to-face format improved by 45.21 percent in fall 2009 and by 48.66 percent in spring 2010, compared with 37.48 percent for online students in fall 2009 and 36.75 percent for online students in spring 2010; the overall mean improvements were 48.56 percent for face-to-face students and 42.97 percent for online students.) When the researchers use ordinary least squares (OLS) regression analysis to determine the significance of delivery format on pre-post-test improvement, in the process controlling for several background variables, they find that delivery format has a significant effect, and that the pre-post test improvement for students in the face-to-face courses is almost one point higher than for students in the online courses. (The coefficient on the online format binary variable is -0.963, which, with a *p*-value of 0.054, is significant at the 10 percent level.)

When the researchers do not control for any background variables, they find that students in the online classes had slightly higher course averages in two of the three semesters (with online students earning mean semester averages of 82.72 and 84.30 for fall 2009 and spring 2010, respectively, and face-to-face students earning mean averages of 81.80 and 77.76 for fall 2009 and spring 2010, respectively), as well as for the three semesters combined (with mean semester averages of 82.18 for the online students and 81.31 for the face-to-face students). In the other semester, spring 2009, the face-to-face students (whose mean semester average was 83.94) outperformed the online students (whose mean average was 79.89). However, when the researchers use OLS regression analysis,

they find the effect of delivery format on course average is not statistically significant. (The coefficient on the online format binary variable is -0.035, which is not significant even at the 10 percent level.)

Finally, Ary and Brune also look at the outcomes by gender (although they do not control for any background variables or use regression analysis, and thus cannot determine whether the differences are significant). In two of the three semesters involved in the study, males in the online format (of whom there were 33 across all three semesters) achieved higher semester averages than males in the face-to-face format (of whom there were 45 across all three semesters). In fall 2009, the mean semester averages were 79.71 for the online format and 78.18 for the face-to-face format, and in spring 2010, the mean semester averages were 81.63 for the online format and 78.17 for the face-to-face format. However, both in the third semester and overall, face-to-face students achieved higher semester averages (84.16 and 77.71 were the mean semester averages in spring 2009 for the face-to-face and online students, respectively, and the overall semester averages were 79.77 for face-to-face students and 79.50 for online students). With respect to pre-post-test improvement, the percentage change between the pre- and post-tests was greater in the face-to-face format for two of the three semesters and overall. (The males in the face-to-face section improved by 40.53 percent and 42.42 percent, in fall 2009 and spring 2010, respectively, and by 41.97 percent overall, compared with 32.61 percent and 23.15 percent for males in the online sections in fall 2009 and spring 2010, respectively, and 37.05 percent overall.) For the third semester, male students in the online format achieved a greater post- pre-test percentage improvement (52.10 percent versus 42.99 percent for spring 2009).

With respect to the female students, those in the online format (of whom there were 58 across all three semesters) had higher course averages than did their counterparts in the face-to-face format (of whom there were a total of 49) in one of the three semesters and also overall; students in the face-to-face format achieved higher course averages than did students in the online format in each of the other two semesters. (The mean course averages for the female online students were 81.41 in spring 2009, 83.99 in fall 2009, 85.84 in spring 2010, and 83.71 overall, compared with mean course averages of 83.83 in spring 2009, 85.18 in fall 2009, 77.07 in spring 2010, and 82.73 overall, for the female face-to-face students.) Female students in the face-to-face format, however, had higher pre-post-test percentage improvements, in each of two semesters and overall; the female online students achieved a higher percentage improvement in the remaining semester. (For female face-to-face students, the mean percentage changes were 53.36 percent for spring 2009, 49.73 percent for fall 2009, 57.56 percent for spring 2010, and 53.00 percent overall. For online female students, the mean percentage changes were 54.88 percent for spring 2009, 39.71 percent for fall 2009, 42.38 percent for spring 2010, and 47.30 percent overall.)

While students' self-selection into one format or the other was not explicitly mentioned in the study, it is unlikely participants were randomly assigned to one of the different formats in light of various (sometimes large) differences in the two groups' ACT scores, pre-course GPAs, and gender compositions. (For example, in the two semesters in which students in the online courses had higher course averages than the face-to-face students, the former also had both higher pre-course GPAs—which Ary and Brune find to be a significant predictor both

of course grades and of pre-post change—and higher average ACT scores.) The comparisons between the face-to-face and online formats that simply report the outcomes without controlling for background variables are thus of limited use.

Moreover, while the researchers' OLS regressions include some controls for background differences between the two groups—such as pre-course GPA, ACT score, gender, semester, major, number of cumulative hours earned prior to semester in which the course was taken, and the difference between pre-course GPA and cumulative GPA at the end of the semester in which the course was taken—there are several other basic background variables which they do not include as controls—for example, race or ethnicity, hours worked during semester, and age. In addition, without randomization, there is always the possibility that other, non-observable differences between the students in the two formats could confound the results.

In addition, some reservations about the course grade comparisons are warranted in light of the different weights accorded to different assignments in the two formats. For instance, the proportion of the course grade based on the final exam was more than twice as high in the online sections (where it was 25 percent of the final grade) as it was in the face-to-face sections (about 11 percent of the final grade). Thus, while Ary and Brune deserve praise for their study's standardization in assessments and assignments across the two formats, and for the set of background data the researchers collected, this study is not without limitations.

Bennett et al. (2007)

Bennett et al. examine the final course grade averages (in the form of percentages) of students who were enrolled in traditional, face-to-face and online versions of principles of microeconomics and macroeconomics courses, taught by three economics professors, at Jacksonville State University in 2005. The study involves a total of 498 students: 406 enrolled in traditional courses and 92 enrolled in online courses. The students' course averages were based largely on their performance on multiple-choice tests, which were identical in the two formats when they were taught by the same professor. The online students and traditional students had the same amount of time to complete the tests, and the tests for the online students were "proctored by university-sanctioned educators."

The researchers hypothesize that, in addition to delivery format, a student's gender, age, GPA, ACT score, and previous economics course experience would affect his or her performance, and compare the traditional and online courses with respect to those characteristics. When the researchers regress students' final course averages on delivery format, the background variables previously identified, binary variables for the different professors, and a binary variable for the microeconomics course, they find no significant difference between the course averages for the different delivery formats when the macroeconomics and microeconomics courses are considered together. However, when Bennett et al. look at student performance only in the microeconomics courses, which contained 188 traditional students and 52 online students, they find that students in the traditional format did better than students in the online format (with final averages of 67.1 percent for the former and 60.2 percent for the latter), and that the difference is significant at the 10 percent level. By contrast, the 40

students in the online macroeconomics classes (who had a final average of 81.2 percent) outperformed the 218 traditional macroeconomics students (who had a final average of 71.6 percent), a difference that is significant at the 1 percent level.

Bennett et al.'s relatively extensive data collection efforts enable them to detect, and then control for, several differences between the online and traditional students taking the courses involved in the study. As a result, they learn, for example, that while women outnumbered men in both the macroeconomics and microeconomics courses, and in both delivery formats, the proportion of women was significantly higher in the online version of the microeconomics classes than in the traditional version of that course. Moreover, in the macroeconomics course, students in the traditional classes had significantly lower GPAs and were significantly more likely to have taken a previous economics course, than were students in the online classes. In both the macroeconomics courses and the microeconomics courses, online students were significantly older than traditional-format students (with average ages of 26.7 for the former versus 22.4 for the latter).

However, while the researchers are wise to identify and control for several background variables that might affect students' course averages, there are a number of other background variables whose inclusion would have improved their analysis. To their credit, the researchers acknowledge the importance of examining other factors that could influence students' economics course performance, such as a student's math background (which might be better accounted for by more specific, math-focused variables than by composite measures of ability such as total ACT score or cumulative GPA), class rank, employment, ethnicity, income level, and what the researchers call "personality type." In addition to the variables identified by the authors, parental education, class year, English proficiency and part-time versus full-time enrollment status would also be useful controls. At the same time, even including a more comprehensive array of background variables still would not eliminate the possibility that the online students and the traditional students differ in unobservable ways; random assignments of students to the different formats would be necessary to prevent unobservable differences in the online and the traditional groups from confounding interpretations of differences in learning outcomes. (While there is no explicit mention of participants' being allowed to self-select into their formats, the existence of several significant differences between students in the two formats suggests self-selection is likely to have occurred.)

Other factors also make it difficult to know the extent to which it is fair to compare the performance of the students in the two formats. The authors indicate that the format of the tests was not the same for all the professors involved in the study: one professor's tests included only multiple-choice questions, while another professor's tests consisted of 60 percent multiple-choice questions and 40 percent "problems" (no further specification is offered); the third professor included a combination of multiple-choice questions and fill-in-the-blank items. It is also unclear what weight tests were given, relative to other assignments (such as problem sets), in calculating students' final course averages, nor do the authors indicate the extent to which the curricula and the content of the course materials and other assignments were different for students in the online format versus in the traditional format. Any differences between the formats in those respects could confound a comparison of the delivery formats.

Finally, given the tremendous variety that exists, generally speaking, with respect to online courses, it would also have been helpful for the researchers to provide more information about the features of those involved in the study, so that the reader could know the types of online courses to which the study's findings about student performance in the different formats are most applicable.

Carpinelli et al. (2006)

Carpinelli and his colleagues look at a sample of 3,491 students taking 150 courses at the New Jersey Institute of Technology over the course of two semesters. Of the 3,491 students and 7,701 course enrollments in the sample, 1,156 students and 1,683 enrollments constitute the “distance learning” (defined only as “eLearning”) group; the other students and enrollments in face-to-face courses constitute a control group for the researchers. The types of files used by the instructors in the distance learning courses varied—some included multimedia files, some included text files, some included both types of files—as did the rank of the instructor and the software platform used (which was generally either WebCT or WebBoard). (The WebCT and WebBoard platforms both allow for online chatting, discussion forums, the sharing and storage of documents, the posting of assignments and quizzes, and the tracking of grades.) The researchers collected background data on student ethnicity, gender, and SAT scores. After comparing the participants' final course grades in the distance learning and face-to-face courses, Carpinelli and his colleagues find that students in distance learning courses earned significantly higher grades than students in face-to-face courses.

Although the sample size of this study is quite large, the design itself has a number of limitations. First and foremost, there is no random assignment of students or attempts to measure or control for differences in the characteristics of the students who enrolled in the face-to-face or distance learning courses. Because the participants in the study self-selected into a particular format, there is no way to know the extent to which the significantly higher grades in the distance learning group are related to the learning delivery format itself, versus the extent to which they are related to the characteristics of the students who enrolled in those courses (for example, the most motivated and talented students might have been attracted to distance learning courses significantly more so than to face-to-face courses).

In addition, there is no mention of an attempt to standardize course materials, course content, or instructor experience between face-to-face and distance learning sections of the same course, which can confound interpretation of the results and make comparing the delivery formats per se difficult. Moreover, while the study mentions variability in the distance learning courses included in the sample—with respect to software platform and presence of multimedia—it does not specify how much variability there is, nor do the authors indicate how much variability there is among the face-to-face courses in the sample. Any conclusions reached from this study concerning the effectiveness of the two formats would, at best, have to be drawn with respect to the effectiveness of distance learning courses, on average, relative to face-to-face courses, on average. While this not a problem as such, it does make it difficult to know to which types of online courses any findings about comparative effectiveness would apply. Finally, the fact that course grades—which often are subjective measures, at least relative to scores on multiple-choice exams—are the only measure of performance used to

compare students in the two formats raises the question of whether instructors' grading standards are comparable in the distance learning and face-to-face courses—a question for which the study provides no answer.

Christou, Dinov & Sanchez (2007)

Christou, Dinov, and Sanchez essentially conducted two studies with a total sample of about 230 students at the University of California, Los Angeles, in both cases comparing a class of students who used the Statistics Online Computational Resource (SOCR) with a class of students who did not use SOCR, and who took the course during a different semester. SOCR (<http://www.socr.ucla.edu/>) is a free, web-based resource that provides data-based educational materials and interactive tools in probability and statistics, such as class notes, practice activities, statistical calculators, interactive graphical user interfaces, computational and simulation applets, and tools for data analysis and visualization. One of the two studies involves a total of approximately 160 students in Stats 13: Statistical Methods for the Life and Health Sciences, across the SOCR-based and non-SOCR-based sections; the other study involves a total of 73 students—39 in the SOCR-based section and 34 students in the non-SOCR-based section—in Stats 100A: Introduction to Probability Theory.

In the study involving Stats 13—which primarily enrolls students who intend to go to medical, graduate, and other professional schools after finishing college—both the SOCR-based section and the non-SOCR-based section received five hours of instruction a week, consisting of three lectures, one discussion, and one lab. Each of the two sections taught in connection with this study, which were each led by two teaching assistants, was split into three sub-sections during the discussions and laboratory sessions. In the lectures for the SOCR-based section, the instructor used SOCR tools for demonstrations and data analysis, and students used the tools to complete projects, labs, and homework assignments. All students were graded according to the same standards. With respect to the study comparing sections of Stats 100A—which attracts mainly mathematics, economics, and computer science majors—the article says that the course met three times a week with the instructor, and once a week for a discussion with a teaching assistant; whether this was true for only one of the sections in the study (the SOCR-based section or the non-SOCR-based section) or for both sections is not clear. In addition, while the abstract of the article states that a “randomized controlled study” was used for these two courses, the description of the study design in the text of the article says nothing about randomly assigning students to one format or the other.

When Christou, Dinov, and Sanchez compare the students' exam grades and “overall performance” across the two formats, controlling for “all possible predictors (like age, major, learning style, background, attitude towards the subject, etc.),” they find that the section of Stats 13 that used SOCR had significantly higher overall performance. Compared with the section of Stats 13 that did not use SOCR, the SOCR-based section did significantly better on two of five tests, and significantly worse on one of the five tests (for the other two tests, there was no significant difference between the two formats). In the Stats 100A study, the SOCR section did significantly better on three exams and in terms of overall performance (for the other two tests, there was no significant

difference between the two formats). For both the Stats 13 study and the Stats 100A study, the article says that the SOCR-based section and the non-SOCR-based section “had the same grading style (exams and homework) and therefore they are comparable.”

However, the fact that the section that students who used SOCR did significantly better overall should not come as a surprise, as the article suggests that students in the SOCR-based sections had as much face-to-face instruction as did the students in the non-SOCR-based sections, plus the benefit of the SOCR resources; one would expect students to do as well, if not better, with more resources than with fewer resources. Another complicating factor is the fact that the non-SOCR based section of the Stat 13 study had a required lab hour in which students used STATA and thereby gained exposure to a type of technology-based instruction.

Finally, even though the abstract to the article describes the research as a “randomized controlled study,” the lack of mention of random assignment of students in the text of the article, as well as the fact that students in the non-SOCR-based sections were taking their courses during a different term than the students in the SOCR-based sections, suggest that random assignment of students did not actually occur. While Christou and his colleagues claim to control for “all possible predictors,” it is unclear how they do so, given that the researchers seem to use two-sample *t*-tests of the means to compare the outcomes of students in the two formats. Thus, it is unclear to what extent any differences observed in the outcomes can be attributed to the different formats versus to differences in the students who enrolled in the different formats.

Colorado Department of Higher Education (2012)

In the first part of this study, the Colorado Department of Higher Education compares the final grades of students in online and face-to-face science courses who were enrolled in the Colorado Community College System (CCCS). The researchers then followed a subset of these students who transferred to four-year public state colleges; in the second part of this study, they examine these students’ performances in science courses at those four-year institutions. Six community college science courses are involved in this study: BIO 111 and BIO 112, CHE 111 and CHE 112, and PHY 111 and PHY 112. For the first part of the study, the researchers compare the course grades of all the students who took these courses in an online format, with the course grades of a random sample of students who took these courses in a face-to-face format. (The sample of face-to-face students is about equal in size to the number of online students in the study.) Students in the study self-selected into one of the two delivery formats. In total, the DHE look at data from 4,585 CCCS students, of whom 2,395 were enrolled in online science courses and 2,190 in traditional-format science courses. The division of the 4,585 participants among the science disciplines is as follows: biology: 1,654 students; chemistry: 1,855 students; and physics: 1,076 students.

In the second part of the study, the researchers examine data for 936 students from the original sample who transferred to one of four public four-year institutions within Colorado. Of these 936 students, 630 had enrolled in at least one biology, chemistry, or physics course at their four-year institutions, the researchers then further limit the sample to the 444 students who had taken either an online course or a traditional course (but not both) at a college within CCCS.

The researchers collected data from fall 2007 through fall 2009 from community college records. For students who then transferred to four-year public institutions within Colorado and took science classes at those four-year colleges, the data from community college records were matched with data from the Student Unit Record Data System. The data collected include cumulative GPA, cumulative credit hours, and grades in the relevant science courses. In the second part of the study, the researchers compute students' average science GPAs at the four-year institutions (in addition to students' biology, chemistry and physics course grades) and then compare students' average science GPAs.

With respect to the first part of the study (concerning grades earned while students were enrolled in community college science courses), students in the traditional-format courses had higher average grades; this difference is significant at the 5 percent level. Looking at the data by subject, course grades were significantly higher (at the 1 percent level) for students in the traditional format in both chemistry and biology courses, but for physics courses there is no significant difference (at the 10 percent level) in grades.

In the second part of the study, when the researchers compare the average science GPAs of a subset of participants at four-year institutions, they find no statistically significant differences by format. There also are no significant differences between the formats when researchers just look at student grades in individual subjects at the four-year institutions.

However, the Colorado Department of Higher Education's study, while praiseworthy for attaining a large sample size, has several limitations. The most critical limitation is the fact that the students in this study self-selected into the format of their choice, leaving open the possibility that students who self-selected into the online format are substantially different from students in who self-selected into the traditional format in ways that could have affected their performance in the relevant science courses. Even in the few variables about which the researchers collected background data from the student record system—cumulative GPA and cumulative credits—differences are evident between the students who self-selected into the online format and students who self-selected into the face-to-face format. For example, among all the 4,585 students enrolled in the CCCS science courses (whose data are analyzed in the first part of the study), students in the online courses have both higher GPAs and higher numbers of credit hours relative to students in the face-to-face courses (the differences are both significant at the 5 percent level). The students in the online classes also have higher numbers of cumulative credits when the data are disaggregated by subject; in the biology courses, the difference is significant at the 10 percent level, and in the chemistry courses and physics courses, it is significant at the 1 percent level. Similarly, in the chemistry courses and physics courses, the online students have significantly higher cumulative GPAs. (This difference is significant at the 1 percent level for chemistry courses and at the 10 percent level for physics courses.)

While the fact that online students generally have higher cumulative GPAs than the face-to-face students may suggest that the online students should do better in the science classes involved in the study, that may not be the case. Cumulative GPA—which reflects student performance in all courses taken up to

that point—is not necessarily predictive of students’ ability to do well in specific science courses, particularly if the courses that the student has previously taken are substantially different from science courses (as might be true of English or history courses, for example). Moreover, even if there is a correlation between cumulative GPA and grades in the science courses involved in the study, merely knowing that students in online courses tend to have higher cumulative GPAs in general would not indicate to what extent any differences observed in the science grades of participants in the different formats can be attributed to differences in cumulative GPA. Had the researchers attempted to control for cumulative GPA in comparing the science grades of students in the two formats, the researchers would be better able to isolate the effects of the delivery format per se on student performance in the science courses.

In addition to gathering data on participants’ cumulative GPAs and credit hours, the researchers could have improved their study by collecting data on other background variables that could be related to student performance. For example, they could have collected information about the income levels of the families of the students, on student race/ethnicity, age, part-time versus full-time enrolment status, gender, and so on. The researchers then could have used regression analysis to compare students’ grades in the two formats and, in the process, controlled for those additional background variables. However, an even better alternative to gathering data on, and then attempting to control for, observable differences between the two populations, would have been randomly assigning students who were in the study to one format or the other (though it is unclear whether this would have been practical in this context).

The researchers also do not supply some “cell sizes” that would be useful to know in interpreting the data. Although the authors provide the total number of students in each of the three subjects, and the total numbers of students in the two formats, it would have been helpful to know, for example, how many of the 1,654 biology students were in the online format versus in the face-to-face format, and how the biology students in each format were divided between BIO 111 and BIO 112. For example, if it were true that 90 percent of the online biology students were in BIO 111 whereas 90 percent of the face-to-face biology students were in BIO 112, it would be unfair to compare the online biology students as a whole to the face-to-face biology students as a whole.

Finally, the researchers do not indicate what efforts, if any, they make to standardize the content, assignments, assessments, reading materials, grading standards, and various other components of the courses across the two delivery formats. If, for example, the online and face-to-face versions of BIO 111 differed dramatically in what assignments counted toward the final course grade and how those assignments were graded, then differences observed in the course grades between the online and face-to-face BIO 111 classes may be more attributable to differences in the assignments given and in the grading of those assignments, than to the format of the course. Similarly, differences in instructor quality that are not accounted for could also confound the comparisons. Thus, one cannot conclude definitively that the differences in course grades seen in this study are indicative of the relative effectiveness of the different delivery formats.

Daymount & Blau (2008)

Daymount and Blau conducted their study at a large public university in a large eastern metropolitan area. Their sample consists of students who were enrolled in seven sections of an undergraduate organization and management course in the fall 2006 and spring 2007 semesters. Two of the seven sections, one in each semester, were taught in an asynchronous online format; together the two online sections enrolled 64 students. The other five sections, which enrolled a total of 181 students, were taught in a “primarily traditional, format with some web-based components.” Participants self-selected into one of the two formats.

The students in the traditional sections met for in-class discussions, in which they spent most of the time “review[ing], explain[ing], or elaborat[ing] on the material in the textbook;” the other class time was used for “learning activities or discussion of other management related issues.” By contrast, the online sections did not have any required face-to-face meetings. While students in the online sections were invited to sit in on traditional class meetings whenever they wanted, the researchers say none did so to the best of their knowledge.

The outcome measures examined in the study include average quiz scores, which took into account students’ performance on 10 quizzes, and final (letter) course grades. In both formats, the final course grades were based mainly on students’ quizzes and on forum discussions; in the face-to-face format, class participation also had a “small influence” on final course grades. The quizzes consisted of, on average, about 30 multiple-choice questions, and open-book quizzes were generally taken through Blackboard; each quiz was made available to students to take for a period of about two weeks. The quizzes collectively constituted about 50 percent of the final course grade in all sections. About 10 percent of the questions on each quiz were identical; the remaining quiz questions were randomly selected from the same question pool.

The forum discussions were also conducted through Blackboard. Instructors initiated about 10 separate discussion threads; students were expected to respond to the threads that interested them the most or to which they thought they could best contribute, with no requirements about to which threads or to how many threads the students needed to respond. Students were graded based on a combination of the quantity and the quality (as determined subjectively by the instructor) of their contributions. Given that students in the online sections did not have any required face-to-face meetings, the standards for earning a given grade in the discussion forum was “approximately twice as high” for students in the online sections as for students in the face-to-face sections. The final grades of students who withdrew from the course (and thus received a “W”) before a predetermined deadline are not analyzed.

The independent variables the researchers examined include student gender, cumulative GPA (excluding the organization and management course involved in the study), the school in which the student was enrolled within the university (if the student was enrolled in the business school, the student’s specific area of concentration within the business school), and the semester being analyzed (fall 2006 or spring 2007).

Daymount and Blau use ordinary least squares regressions to analyze the data. When they regress each outcome variable on delivery format only, they find that there is no significant difference between the students in the two formats with respect to final course grade or with respect to tendency to withdraw from the course before a grade was assigned. With respect to final course grade, students in the online sections did slightly better, scoring, on average, about one-quarter of a letter grade higher than the traditional-section students—a difference that was significant at the 5 percent level. When the researchers include as controls the independent variables identified in the previous paragraph, they find that the difference in final course grades remains insignificant and decreases (to 5 percent of a letter grade, when all the controls are added). This finding, combined with the facts that the students in the online sections had higher GPAs, and that GPA is a significant predictor of final grade, suggest that any difference between the final course grades earned by the students in the two formats more likely reflects pre-existing differences between the populations of students who self-selected into each format. Similarly, the difference between the two formats with respect to average quiz score, which is significant in the absence of controls, decreases and loses its significance when the controls are added to the regression. (Adding controls decreases the size of the coefficient from 2.84 points and a 5 percent significance level to 1.18 points and no statistical significance.) Given that GPA is also a statistically significant predictor of average quiz score, differences in GPA between the two formats also are likely to account for differences in average quiz score.

As Daymount and Blau note, there are a number of significant differences between the students in the face-to-face and online groups. For example, in addition to having significantly higher cumulative GPAs, the students who took the online format are significantly more likely to be female, significantly more likely to be seniors, and significantly less likely to be transfer students. It is quite conceivable that the students in the two different formats differ in other ways than in the variables for which data were collected in the study, both with respect to observable characteristics, such as race, age, and income level, and also with respect to unobservable characteristics. Although data can be collected on observable variables, one cannot ensure that the students who enroll in the different formats are not different in unobservable ways without randomly assigning students to one format or the other. Thus, while the researchers should be commended for controlling for a number of background variables, there are others which they fail to control for; randomly assigning participants would have resulted in a still better study design than the quasi-experimental one the researchers used. Similarly, Daymount and Blount deserve credit for their attempts to standardize the quiz questions across the two formats, though it also should be noted that differences in instructor quality may still confound the results. Finally, while the final course grades for students in both the online and face-to-face formats were based primarily (if not exclusively) on quiz scores and forum discussion responses, holding the online students to a standard that was twice as high with respect to the latter component may have unduly biased those students' scores downward; even if one is sympathetic to the rationale for that method of grading, it is still probably unfair to compare the discussion forum performance of students in the two formats using such different standards.

Dell, Low & Wilker (2010)

Dell and his colleagues compare an online educational psychology course with a face-to-face version of the same course. Of the 59 undergraduates who are involved in this study (the relevant population given the inclusion criteria for this literature review), 13 were enrolled in the online section of the course, and 23 were enrolled in each of two face-to-face sections of the same course. Of the 46 students in those two face-to-face sections, 13 were male and 33 were female, compared with four males and nine females in the online section. The online course included a forum for asynchronous group discussions; the instructor played a limited role, occasionally checking in on the discussions, providing limited feedback, clarifying misunderstandings, and occasionally asking questions to prompt critical thinking. The same instructor taught all three sections, using the same assignments, readings, activities, and discussions in the same order. All online students took the final exam with a pre-approved proctor.

The researchers compare student performance on three assignments: a two-to three-page report on an experiment conducted by the student (which was followed by an online or face-to-face discussion of the experiment outcomes); the student's analysis of a learning and teaching episode observed after reading in the textbook about the concept of a zone of proximal development; and an essay-format question about self-efficacy on the final exam. The latter was graded by "independent raters in a blind review" according to a rubric, and the raters were required to reach an agreement on all ratings.

Dell and his colleagues find that, on average, students in the online group did significantly better with respect to the report on the experiment (the difference is significant at the 10 percent level), but that there is no significant difference in the two groups' performance with respect to the analysis of a learning episode or the essay-format question on self-efficacy.

Like other studies in this review, this study, by itself, does not hold very much weight. Not only is the sample size relatively small, but the study design is not very rigorous, with no random assignment of students to the online format or face-to-face format, and no attempts to measure or control for differences in the populations of students who self-selected into the different formats.

Dinov & Sanchez (2006)

At the University of California, Los Angeles, Dinov and Sanchez compare the performance of 20 students in the hybrid version of a course called Introduction to Probability, to that of 39 students in a face-to-face version of that same class. The students in the hybrid group, but not the students in the face-to-face group, used modules of SOCR (see description in the Christou et al. study above) and SOCR applets to solve problems and to complete homework assignments. Both groups received instruction three mornings a week, and attended one discussion section per week. Students in both formats were taught using the same PowerPoint lectures (which were also available on the course website), followed the same syllabus, used the same textbook, and participated in similar discussion sessions. They were also given the same practice midterm and final exams, as well as the same homework assignments, which were graded using the same rubric; however, only the hybrid class was required to use SOCR in some of the homework problems (on which the face-to-face class was required

to use the statistical software package R). The hybrid class was also assigned additional activities in homework assignments that required SOCR. The two groups' exams were graded at the same time, with the same instructor grading the same few questions on all students' exams. The students in the face-to-face class and hybrid class were "as similar as possible in all characteristics relevant to the outcome measures of interest," except that the face-to-face class had more graduate students than did the hybrid class (sixteen graduate students in the face-to-face class versus three graduate students in the hybrid class). In this study, the researchers compare the homework, midterm, and final exam scores and the overall course grades of students in the two classes.

The researchers find that the undergraduate students in the hybrid group did slightly better on all measures than did the undergraduate students in the face-to-face group, though they do not indicate whether these differences are statistically significant. The variance in scores is also much lower in the hybrid group than in the face-to-face group for all performance outcomes. However, when graduate students are included in the analysis, the face-to-face group outperformed the hybrid group (presumably with respect to all outcomes, although the study does not specify this).

Like the other studies, this one also has its limitations. There is no mention of random assignment of participants to the hybrid group or the face-to-face group, and while the authors provide the composition of each group by class year and by major, there does not appear to be any attempt to control for any of the differences between the two formats, with the exception of the fact that the authors show the results in both groups with the graduate students excluded. The authors do not show how the face-to-face group and the hybrid group compare with respect to other demographic variables, such as age, race, or gender, or with respect to measures indicative of academic preparedness, such as GPA. Finally, the hybrid group had a fair amount of face-to-face instruction—seemingly as much as the face-to-face group. In fact, the only apparent differences between the two formats are that the hybrid group had SOCR-based activities in addition to what the face-to-face group was given, and that the face-to-face group (but not the hybrid group) was required to use the R software for some assignments. Thus, it should not come as a surprise that a class that received instruction three times a week, attended a discussion once a week, and had additional SOCR-based activities, should outperform a group that received instruction three times a week and attended a discussion once a week, but was not given those additional activities.

Driscoll et al. (2012)

Driscoll and his colleagues examine an undergraduate introductory course that is cross-listed in the sociology and anthropology departments at a "large urban university" in North Carolina. Although the researchers do not provide the name of the course, they state that it concerns "the role of technology in societies throughout history and in processes of societal change," and that the course is open to all students, with no prerequisites, and can be taken to satisfy one of the university's general education requirements. The researchers examine the performance of students from six sections of the course over three consecutive terms: spring, summer, and fall of 2010. One online section of the course was offered during each of the three terms; one face-to-face section was offered

during the spring term, and two face-to-face sections were offered during the fall term. The same instructor taught all six sections, each of which was capped at 80 students. Although a total of 231 students were enrolled in the three face-to-face sections, and a total of 212 students were enrolled in the three online sections, not all of those students consented to be in the study; instead, the researchers end up with a sample of 368 participants, 170 of whom are online students and 198 of whom are face-to-face students. (The sample size is based on the number of students in the study sections who completed a survey administered at the end of the semester, and who provided informed consent and the last four digits of their student ID numbers in order to link students' survey responses to their grades.)

The face-to-face version of the class consisted of in-class lectures and discussion sessions two to three times a week, complemented by the use of some forms of technology, including e-mail, online assessments, and readings posted on the learning management system (LMS). Students in the online version of the course learned the material through PowerPoint lecture slideshows and other course content that was posted in the LMS (which was the same for both formats). The online students communicated with each other and with the instructor via asynchronous, threaded discussion boards, which were also available through the LMS. Students were required to respond to certain prompts and to each other's posts; the instructor graded and offered feedback on students' posts, as well as raised additional questions. The instructor further communicated with students by posting announcements on the LMS, which were also e-mailed to students, to "clarify instructions," "remind students of upcoming deadlines," and "provide broad feedback on the classes' performance and grasp of course concepts."

In the spring and fall terms, during which both online and face-to-face sections were taught, the learning goals and outcomes, content, readings, and standards were identical for both the face-to-face and online sections taught during that term. In general, all sections involved in the study "largely covered the same course material"; that being said, "some minor refinements were made from term to term [over the three terms involved in the study], but the books, supplemental readings, and section topics were largely consistent throughout." The material was presented in the same order in the online and face-to-face versions of the course and on the same schedule. "By constructing the online sections in this fashion," the researchers write, "the instructor was able to make the online and F2F [face-to-face] sections as similar as possible while designing the online sections according to pedagogically based teaching practices."

To measure students' academic performance, the researchers look at students' scores on one unit exam—which included a combination of multiple choice, short answer, and essay questions—and on an "integrating data analysis" (IDA) assignment—in which students were asked to find and analyze "various national-level indicators of development, well-being, and technological advancement," and to discuss the effects of technological innovation and the unequal distribution of its benefits. The examination was administered through the LMS; students were given a four-day window to complete the exam. The exam, the IDA assignments, and the rubrics for grading those evaluations were the same for the students in the two formats. This exam covered a unit of the course whose content did not

change over the course of the three terms involved in the study. Students' scores were on a 0-to-100-percentage-point scale and were made available to researchers from the instructor's records of students' grades on these assessments.

Students' responses to a survey administered at the end of the semester also provided background data about student age, gender, GPA (which researchers round to the nearest half-point increment, yielding seven possible values between 1.0 and 4.0, inclusive), the number of credits in which the student was enrolled at the time he took the course involved in the study, the number of hours per week the student worked for pay, class year (which researchers use to create a binary variable indicating whether the student was either a senior or a non-senior, as "preliminary analysis revealed that this was the only distinction that affected student performance"), and the number of online courses taken previously.

Driscoll and his colleagues analyze the data using Ordinary Least Squares (OLS) regression. When the researchers do not control for any background variables, they find that students in the online format performed significantly worse on the unit exam (a difference, significant at the 1 percent level, of almost three points on the exam) than students in the face-to-face format. However, when they introduce student GPA (which is significantly and positively correlated with exam performance) as a covariate into their regressions, the difference between the two formats in exam performance is halved and is no longer significant at the 5 percent level, the lowest level of significance used by the researchers. This finding reveals a selection effect: students who enrolled in the online sections tend to have significantly lower GPAs (and perhaps are generally weaker students) than students who enrolled in the face-to-face sections. Several other independent variables, including class year, number of online courses previously taken, credit hours completed, gender, and hours worked for pay per week, each have an insignificant effect on performance, and do not affect the absence of a significant effect of delivery format on performance. Comparisons of face-to-face and online students' performance on the IDA assignment also suggest that different types of students were self-selecting into the different formats. When no covariates are included, online students again did significantly worse on these assignments than face-to-face students, with a difference, significant at the 1 percent level, of about 2.5 points; however, when student GPA is included as a covariate, the difference drops to about 1.4 points and is no longer significant at the 5 percent level. As with students' performance on the unit exam, the effects of delivery format on students' performance on the IDA are not affected by the addition of other covariates.

Compared with much of the other research in this literature review, this study is well-done in terms of including a number of covariates and using a fairly large sample size. In addition, the fact that the same instructor taught all study sections, combined with the fact that a number of components of the classes in the two different formats were standardized across the six sections, limit the extent to which differences in curriculum or assignments confound the researchers' ability to compare the effectiveness of the two delivery formats. At the same time, one must note that the fact that participants were able to enroll in the format of their choice raises questions about the extent to which Driscoll and his colleagues are able to isolate the effects of delivery format per se from differences in the populations of students who self-selected into those formats. This caveat is reinforced by the differences the researchers observe between the

students in the face-to-face sections and the students in the online sections. The researchers write, “The students in our study who selected an online section of the course tended to be older, tended to have lower GPAs, tended to have greater experience with online courses, were more likely to be college seniors, were taking fewer credit hours, and were working a greater number of hours per week. These factors indicate a substantially different composition of the student population who take online courses compared with those who take F2F.” While the researchers are able to control for the differences they observe with respect to these particular variables, the face-to-face and online students may also differ with respect to variables other than those the researchers attempted to control for, such as income level or race/ethnicity, as well as unobservable (or less easily observable) differences. A randomized design would have been preferable in order to ensure that the populations in the two formats are as comparable as possible—as the researchers, to their credit, acknowledge: “Had we been able to randomly assign students to the online or F2F sections. . . , the possibility of a selection effect would have been eliminated and we would have been able to more directly observe the influence of the classroom setting.”

Enriquez (2010)

Enriquez looks at a sample of students enrolled in an introductory, three-unit course in circuits analysis, which is required for all engineering majors, over two semesters at Cañada College. A member of the California Community Colleges, Cañada College has approximately 6,000 students and a sizeable Latino population (comprising about two-fifths of the student body). Over the course of the two semesters, a total of 25 students enrolled in the online version of the course and 30 students enrolled in the face-to-face version of the course. Students in the online sections logged into an online classroom several times a week for 15 weeks, at the same time face-to-face students were attending lectures. The online students experienced the lectures through a combination of Tablet PC (which replaced Blackboard and was used to generate lecture notes) and Elluminate Live! video-conferencing software (which allows for web-based synchronous delivery). Students accessed the course through the eCollege delivery platform, which included content tools for viewing materials and assignments, and submitted their homework through Moodle, an open-source learning management system that was also used by the face-to-face students in this study. The online students did not attend any face-to-face lectures during the semester; they only needed to come to campus to take four tests and the final exam. Students in both the face-to-face and online sections were given identical tests and homework problems (with the same deadlines).

Compared to the students taking the online sections of the course, a greater proportion of students in the face-to-face sections were electrical engineering majors (30 percent versus 16 percent), and a greater proportion of face-to-face students were taking the lab course, which was designed to reinforce concepts learned in lecture, concurrently with the lecture course. According to the author of the study, the latter difference gave the face-to-face students an “advantage over online students in that they’re able to apply and experimentally verify concepts learned in the lecture course.”

At the end of the semester, Enriquez compares the scores of students in the two groups on fifteen homework assignments, four tests, and the final exam, as well as the proportion of students who completed the course and the proportion of students who earned an A, B, or C in the course. He finds that students in the online sections earned slightly higher course grades and scored slightly higher on their homework assignments and on the final exam, but did slightly worse on the other tests, though these differences are not statistically significant. In addition, the retention rate and the percentage of students who received an A, B, or C are equivalent across the two groups.

However, while the sample sizes in this study are not extremely small, they also are not large, nor is there any mention of randomly assigning students to one format or the other. In addition, all of the online resources available through Moodle were available to both the face-to-face students and the online students, so it is unclear how much difference there actually was between the two formats. The study also mentions a computer projector that “on-campus” students used to view lecture notes, suggesting that the students in the face-to-face format might actually have been experiencing the lecture through a broadcast/telecast rather than face-to-face. Finally, while the author identifies some differences in student populations in the different formats, he makes no effort to control for them in his analysis of the outcomes.

Friday et al. (2006)

This study examines student performance in two undergraduate courses at a commuter university in a major metropolitan area in the southeastern United States. These courses, both in the school’s College of Business Administration, are entitled Organization and Management and Strategic Management; the former is a required introductory course for all business majors, while the latter is a capstone core curriculum course which is also required for all business students and which students typically take during their final semester at the university. Eight semesters are examined in the study; during this time period, 213 students registered in traditional sections and 380 were enrolled in online sections of the Organization and Management course, and 456 students were registered in traditional sections and 298 students were registered in online sections of the Strategic Management course. Students were allowed to self-select into the section of their choice for both courses.

In each of the two management courses, students learned the same content and took the same exams in both the online and face-to-face formats. The face-to-face section involved student presentations, guest speakers, and videos, and the instructor kept track of students’ attendance and tardiness. Students in the online sections were required to post summaries of management-related articles to the “Public Bulletin Board.” (The attendance and tardiness of students in the online sections were not tracked, and students in the face-to-face sections did not have to post article summaries to a bulletin board.) The same instructor taught both the online and face-to-face sections of both courses during the spring, summer, and fall semesters of 1999 and 2000 and during the spring and summer semesters of 2001.

The outcome of interest in this study is final course grade, which was based on a student's scores on assignments, exams, and group projects and presentations, and on his participation and attendance (for the face-to-face sections). The researchers gathered data on student gender, but not on any other demographic variables.

In addition to delivery format and gender, the researchers' analysis takes into consideration the instructor's experience and the semester in which a student was taking the course; it also includes an interaction term for delivery format and gender. The researchers analyze separately the eight semesters' worth of data for the Organization and Management course and the eight semesters' worth of data for the Strategic Management course. An analysis of co-variance in the Organization and Management course data shows that students in the traditional sections earned, on average, higher grades than students in the online sections, with mean grades of 8.12 and 7.35, respectively, when the grades are converted into numbers (with an A being equivalent to a 12, an A- being equivalent to an 11, and so on) and no controls are used, although the significance (if any) of this difference is unclear. However, there is no significant difference (even at the 10 percent level) in students' grades in the two different delivery formats when researchers control for instructor experience and for the semester in which the student was taking the class; there also are no significant interaction effects between course delivery and gender.

With respect to the Strategic Management course, the students in the traditional sections similarly earned higher grades than students in the online sections when no controls are used, with students in the traditional sections earning, on average, the equivalent of an 8.71 and students in the online sections earning, on average, the equivalent of an 8.38. As is also true in the Organization and Management course, students in the Strategic Management course did not perform significantly better in one format or the other when researchers control for instructor experience and semester of enrollment. However, unlike in the Organization and Management course, there is a significant interaction between delivery format and gender in the Strategic Management course, with men in the online format receiving, on average, lower grades than men in the traditional format (with mean grades of 8.05 and 8.70, respectively). (Men in the online format also earned also lower grades than women in both formats, whose average grades are 8.71 for women in the online version of the course and 8.73 for women in the traditional version.)

Although the large sample sizes and the standardization of course content and exams are commendable, this study has a few drawbacks. The primary deficiency (which is by no means unique to this study) is that students were allowed to self-select into the format of their choice, leaving open the possibility that, in each of the two management courses, students who self-selected into the face-to-face format are substantially different from students who self-selected into the online format in ways related to their performance in those courses. Furthermore, while the researchers implicitly acknowledge that the students who self-selected into online courses may differ from the students who self-selected into the face-to-face versions of those same courses, they only control for possible differences with respect to student gender. Controlling only for differences in gender is problematic, as the two populations are likely to differ with respect to other demographic variables that could also be related to course performance, including (but not limited to) race/ethnicity, age, part-time versus full-time

enrollment status, income level, parental education, or previous academic ability (as indicated by measures such as SAT scores or GPAs). This being said, the researchers, to their credit, recognize that “the limited number of variables in this study precluded the examination of the impact of other potential influencing factors, such as student learning curve, student participation, pedagogical methods, and other professor and student-related outcomes.”

Moreover, the outcome measure used to compare students’ performance also has some limitations. The final course grade is the only outcome measure, and it was not calculated in the same way in the online sections and in the face-to-face sections; attendance was factored into the final course grades of students in the face-to-face sections but not into the final grades of students in the online sections. While the authors say that final course grade was “a composite scoring of the grades students earned on assignments, exams, group projects and presentations, and attendance and participation,” they do not indicate whether, apart from attendance, the weighting of the various other components was the same in the two formats. In fact, the researchers acknowledge that “there was some variance in the way the traditional and online course grades were calculated,” thus reducing the comparability of the course grades of the students in the different formats.

Holdhusen (2009)

Holdhusen compares the outcomes of 14 first- or second-year undergraduate engineering students in a face-to-face engineering graphics course with those of students in three different types of technology-enhanced formats of that same course. Of the 53 students enrolled in technology-enhanced formats, 20 students enrolled in a section that used audiographics technology; 10 students enrolled in a section employing asynchronous online instruction; and 23 students took the engineering course in a hybrid format which combined face-to-face and synchronous online instruction. In the asynchronous online section, instruction was delivered via a learning management system, and included text, images, web videos to demonstrate drawing techniques, and a discussion board for students to post questions and communicate with each other and with the instructor. Students submitted their assignments electronically. Students in the synchronous audiographics mode met face-to-face twice a week for 120 minutes each time; of the 120 minutes in each face-to-face meeting, 20 to 30 minutes were spent presenting new material via web-conferencing software, followed by a virtual lab, similar to a chat room, in which students could ask the instructor questions. The course web site used by students in the synchronous section was similar to the website used by students in the face-to-face section. Finally, the hybrid section met once per week for 150 minutes, during which time students and the instructor discussed the material briefly and students could ask the instructor questions about their assignments. Students in the hybrid section also had the option of chatting online with their instructor for 150 minutes once a week, and used a course web site that was similar to the site used by the purely online section. The lab exercises given to students in each of the four sections were said to be “essentially identical,” and the same instructor taught all sections. Students enrolled in a particular section “based on what was offered to them,” where “[t]he different modes were offered based on the geographical location of the instructor as well as the students.”

When Holdhusen compares student performance on a post-test consisting of 25 multiple-choice questions, which he himself developed and administered to all students at the end of the course, he finds that students in the audiographics section scored the highest, while students in the asynchronous online section scored the lowest. However, the only difference that is statistically significant between two sections is appreciably better performance by the face-to-face section relative to the hybrid section on one topic. The average overall score of face-to-face students on the post-test is not significantly different from the average overall score of students in any other section (though on one particular post-test topic, the audiographics section and the face-to-face section each did significantly better than the hybrid section). When Holdhusen compares the performance of students on very similar lab exercises that were used to measure students' ability to apply engineering graphics concepts to computer-aided design software, he finds no significant differences between any two of the four formats in the total lab scores, though the sub-scores for students in the audiographics section are significantly higher than are the sub-scores of students in the face-to-face section on one particular topic.

The results of this study should be interpreted with great caution. The number of students in each condition is extremely small; the largest section has only 23 students. While it is unclear how much choice the students had in choosing in which section to enroll, it is extremely unlikely the students enrolled in sections completely randomly; there are likely to be differences between the students enrolled in one format and the students enrolled in the other formats. Furthermore, the author notes that there was differential attrition across the sections, and the attrition rate was especially high in the online section, in which about 40 percent of the students who were originally enrolled withdrew or stopped participating, compared with a 10 percent to 15 percent attrition rate in the other sections; the reason for this discrepancy is unclear. Another fact to be wary of is that students in the hybrid and synchronous audiographics sections had a fair amount (at least two hours per week) of face-to-face contact with their instructor, raising the question of how different these formats were from the face-to-face comparison format. For instance, both the face-to-face and the synchronous audiographics sections had course websites and had two 120-minute meetings each week, consisting of the presentation of new material by the instructor, followed by time spent in a computer lab.

Lewis & Harrison (2012)

Lewis and Harrison conduct a quasi-experimental study using two concurrent sections of what they describe only as an "introductory social science course" at a small liberal arts college. The total sample size of this study is 60 students, with 32 in the treatment section, which was taught in a hybrid format, and 28 in the control section, taught in a purely face-to-face format. The treatment section met three times a week for 50 minutes each time. Students in the treatment group were instructed to listen to prerecorded lectures, which were embedded into PowerPoint slides and posted online on the course management system, before coming to class. The treatment section's class time was used for answering questions about the lectures, discussing difficult or unclear points, and participating in interactive group work that synthesized the assigned material. By contrast, the students in the control section met twice a week for 1 hour and

15 minutes each time. Students received material via a passive learning format (listening to lectures) supplemented with the PowerPoint slides; questions were “encouraged.” After the course material had been covered, the same PowerPoint slides with embedded lectures were posted online in the course management system. If there was time left over at end of each unit, the instructor assigned interactive group work.

Several measures were taken to make sure the control and treatment sections were as comparable as possible. Both the control and treatment sections were taught by the same instructor, met face-to-face at the same time of day and in the same classroom (though presumably on different days of the week), met for the same amount of total face-to-face time each week, moved at the same pace through the same textbook, and received the material using the same lecture script and the same PowerPoint slides emphasizing main lecture points. Moreover, both classes were asked to read the assigned chapter in the textbook prior to coming to the class in which it was to be discussed, and the same quizzes, practice exams, and interactive review sessions were posted on the course management web site. A 50-question pre-test, given to both classes on the first day of class to test for equality of potential, revealed mean scores of 27.03 for the hybrid section and 24.46 for the face-to-face section and a *p*-value that is greater than 0.096.

The researchers compare the performance of the students in the two sections using several identical, objective assessments, which were administered in the two sections during the same weeks of the semester. These assessments consisted of two 20-item multiple-choice, in-class quizzes; three cumulative 100-item, multiple-choice, in-class exams; and a cumulative 200-item, multiple-choice final exam, within which a post-test was embedded.

As expected, both sections show significant increases in scores from pre-test to post-test, although there is no significant difference (at the 10 percent level) between the scores of the control and experimental classes on the post-test. (The pre-post test difference for the hybrid students has a *t*-score of 9.77 and the pre-post test difference for the face-to-face students has a *t*-score of 10.14; both *t*-scores are significant at the 0.1 percent level. On the post-test, the hybrid students had a mean score of 43.72 and the face-to-face section had a mean score of 40.82; the *p*-value for the comparison is greater than 0.196.) With respect to performance on the whole final exam (not just the portion of the final exam that constituted the post-test), the hybrid section did significantly better (at the 1 percent level), earning an average score of 168.72, compared with the control group’s 145.83. The hybrid group also scored significantly higher with respect to total course points (which excluded pre-test and post-test scores), and did significantly better with respect to scores on both quizzes and on one of the three exams. (The comparison was 459.18 to 405.89, significant at the 1 percent level, for total course points; 14.03 to 10.56, significant at the 0.1 percent level, for quiz 1; 17.11 to 14.92, significant at the 5 percent level, for quiz 2; and 84.34 to 72.88, significant at the 1 percent level, for the exam.) On the other two exams, there is no significant difference (at the 10 percent level) in the scores of the two groups. (On one of the two exams, the hybrid group scored 76.63 on average and the face-to-face group scored 72.00 on average; on the other exam, the hybrid group scored 81.89 on average, and the face-to-face group scored 75.57 on average.)

However, this study has a fair number of limitations such that the results need to be interpreted with caution. First, the study does not involve random assignment of participants to the control or experimental group. The pre-test results show that there are at least some differences in the students who self-selected into the different formats, and while the difference in mean pre-test scores is relatively small and may be significant only at the 10 percent level (the authors say that the *p*-value is less than 0.096), pre-test scores are only one dimension, and the researchers cannot say for sure that the treatment group and control group do not differ in other, non-observable ways, and thus cannot eliminate alternative explanations for different performance between the two groups. (To their credit, the researchers acknowledge as much, even offering possible alternative explanations: “It could be that the more motivated students sign up for courses that meet more frequently”—i.e. three days a week, as was true of the treatment group, rather than two days a week—or “courses meeting 3 days per week with a shorter course period can sustain students’ attention better than courses that meet twice per week for a longer course period.”)

Moreover, while standardization between the control and the treatment groups generally is praiseworthy and helpful in isolating and identifying the reason for any difference in performance, the fact that the control and treatment groups met for the same amount of face-to-face time each week and had access to almost entirely the same materials posted online in the course management system raises the question of what exactly the treatment was. The comparison involved in the study seems to be less between a face-to-face course, in which a large majority of the instruction was conveyed in class or via textbook, and a hybrid course, involving substantially less face-to-face time and a substitution away from face-to-face time and toward greater use of online course materials, than it is a comparison between one hybrid course that divided a given quantity of face-to-face time each week among two sessions, and another hybrid course that divided the same quantity of face-to-face among three sessions. The relatively small difference between the control and treatment formats limits the study’s usefulness to this literature review, which is concerned with the former of the two comparisons described in the previous sentence. However, the comparison used in this study may make more sense for the researchers’ own purposes and thus cannot necessarily (fairly) be labeled as a “limitation” per se. Lewis and Harrison are interested in “whether or not online delivery assists in promoting best practices in education”—which, from their own literature review, includes “promoting active learning, cooperation, and student-faculty contact”—and they “hypothesized that online delivery of lecture prior to course meetings allows more in-class time to achieve these goals.” Thus, having comparable amounts of face-to-face time in the two groups may allow for them to better test their hypothesis, even if this aspect makes the study less helpful for answering the question with which this particular literature review is concerned.

Mentzer, Cryan & Teclehaimanot (2007)

At a medium-sized campus (enrolling approximately 20,000 students), Mentzer and his colleagues randomly assigned 36 students, all of whom were enrolled in a bachelor’s degree program within the school’s College of Education, to a face-to-face or online section of an entry-level survey course. The course, titled Early Childhood Education: Philosophy and Practice, is required for first-year

students in the teacher education licensure program and typically enrolls about 100 students. Of the 36 participants, 17 females and 1 male were assigned to each group. All students were required to contact the department office before they were allowed to register for the course; once they contacted the department office, they were asked if they wanted to be in the study, and those who agreed to participate were randomly assigned to the face-to-face or online format. (Those who declined were not randomly assigned and could register for sections separate from those of the study participants.) The online students participated in two hour-long “Live Chat” sessions each week in lieu of the class discussions that the face-to-face students attended, and they were assigned to small groups of four to six students each, to complete assignments over the Internet through chat rooms. Both the online and face-to-face sections involved in the study were taught by the same instructor, who used the same syllabi and gave the same group assignments (which took place in class for students in the face-to-face group). Students in the online section did not have face-to-face contact with the instructor during semester, and were not in other courses taught by the instructor. A diagnostic instrument given to all participants at the beginning of the semester showed no significant differences between the face-to-face and online sections in terms of the learning preferences of the students in the two groups.

When the researchers compare the midterm and final exam grades, and the overall course grades, of the students in the two formats, they find that exam performance is not significantly different between the two conditions. However, final course grades are significantly lower for students in the online group (an average of a B for online students, versus an A- for face-to-face students) because the students in the online group were less likely to complete assignments that counted toward their final course grades.

While this is one of the few studies reviewed that uses random assignment of students to an online group or face-to-face group, it is difficult to draw any firm conclusions from the results, mainly because the sample size is quite small. In addition, while the researchers specify that the gender composition and the learning preferences are the same or very similar in the two formats, it is unclear whether they control for these variables or for other possible predictors of performance (for instance, high school GPA for first-year students or family’s income level) in their analysis (which, while not essential for a randomized study, would still be useful for the reader to know).

Olson (2002)

In this study, Olson looks at the performance of students enrolled in online and face-to-face formats of CS 161: Introduction to Computer Science at Western Oregon University. This class is the first course students pursuing a major or minor in computer science are required to take; since the course also fulfills the science component of the school’s general education requirements, it attracts a wide range of students and has the largest enrollment of all courses in the computer science department. Students do not learn to program in this course; instead, the curriculum covers “simple architecture, binary representation of numbers, simple machine language instructions and an examination of elementary data structures and algorithms.” Two or three sections of the course are generally offered each semester, with about 70 students per section. In the

term in which the study was conducted, spring 2001, 140 students, divided into two course sections, were enrolled in CS 161, only four of whom were declared computer science majors.

Olson invited the 140 students enrolled in CS 161 during the spring 2001 quarter to participate in the study; 48 students agreed. These 48 students were randomly assigned either to an experimental group, whose students took the online version of the course, or to a control group, whose students took the course in a face-to-face format. (Olson refers to the latter group as consisting of students “who had volunteered to take the course over the Internet, but were not chosen to receive Internet instruction,” suggesting that the recruiting efforts for the study involved advertising the study as an opportunity to take an Internet-based course.) A third group consisted of students who took the course in a face-to-face format but who had not volunteered to participate in the study—students who, according to Olson, “prefer a traditional lecture presentation.”

The face-to-face version of this course had three hours of lectures each week; the course ran for an 11-week term. The online version of CS 161 involved in this study, which was developed at WOU, consisted of a website that contained general information and downloadable reading assignments, timelines, and detailed PowerPoint slides. On the slides, the material was divided into, and presented in, small chunks in order “to simulate how a professor might show the process one line at a time during a lecture using a white board.” Students in the online section accessed most, if not all, of the written course content online.

The outcomes of interest to Olson are students’ completion rates, as well as a performance-based measure of student “success” in the course. In the all three groups, a student’s “success” is based on the mean of the student’s scores on eight multiple-choice quizzes administered over the course of the semester. There were also multiple forms of each quiz, and each quiz consisted of 15 questions randomly selected from a test bank. In the face-to-face sections, the quizzes were administered on Fridays and covered material presented in the most recent lecture; students in the online format were given one additional week to complete a quiz covering the same material.

At the end of the term, Olson has completion data for 17 students in the experimental group, 18 students in the control group, and 78 students in the third, “traditional” group. (The article says Olson ends up with a total sample size of 111 students; in light of the total number of students for whom he has completion data, it is unclear why this number is 111 rather than 113.) Olson has “success” score data for 9 students in the experimental group, 11 students in the control group, and 55 students in the “traditional” group. (The small sample sizes for the success score analysis reflect the fact that Olson only examines the scores of students who completed the course.)

When Olson examines the differences in completion rates across the three groups, he finds no significant difference at the 5 percent level, using a one-tailed *t*-test, between the completion rate of students in the experimental group (those who were randomly assigned to the online version of the course) and that of students in the control group (who were randomly assigned to the face-to-face version of the course). The completion rate of the students in the “traditional” group (the students who took the course in a face-to-face format but who did

not agree to be randomly assigned) also does not differ significantly from the completion rate of students in the experimental group, or from that of students in the control group. In examining the success scores of students who completed the course, Olson finds no significant difference at the 5 percent level, again using a one-tailed *t*-test, between the scores of students in the control group and those of students in the treatment group. However, the students in the traditional group (who did not agree to be in the study) have significantly higher success scores than do both the students in the experimental group and the students in the control group.

The study design is fairly rigorous as a result of its random assignment procedure and its standardization of assessments between the control and experimental groups. The main deficiency of this study is its small sample size. While 48 students consented to be in the study and were randomly assigned, Olson only has completion data for 35 of those students, and success score data for 20 of those students. These small sizes make it difficult to reach well-founded conclusions about the relative effectiveness of the different delivery modes. Olson also does not mention what happened to the 13 students who agreed to participate but for whom he does not have completion data. If the reason their data are not included is that these students opted not to participate in the study after learning their random assignments, then there could be a selection bias distorting the results of the study; the students who opted not to be in the study after learning their random assignments could be different, in ways related to their academic performance, from the students who learned of their random assignment but remained in the study. Olson also does not specify whether the students in the control and experimental groups had the same instructor; if not, differences in instructor quality could bias the comparison.

Finally, while this may not be a drawback per se, it would have been helpful if Olson had collected some background data on both the students who agreed to participate in the study and those who had not agreed to participate. The fact that students in experimental group do significantly worse than students in the “traditional” group (who did not agree to participate in the study) but not significantly differently from students in the control group (who volunteered to participate), suggests that the students who volunteered for the study are not necessarily representative of all the students enrolled in the course. As Olson states, the finding of no significant difference between the experimental and control groups only “indicates that students preferring to take an online course do equally well in an online course or a lecture setting.” He rightly points out that the fact that participants as a whole did worse than nonparticipants raises the question of “Which student characteristics result in lower scores in the course?” Acknowledging that this question “calls for speculation,” Olson notes that “a large number” of students in the course are not in the computer science department and are likely to be taking CS 161 to fulfill their science general education requirement. He hypothesizes that “students wanting to take the path of least resistance” might have been especially likely to volunteer to sign up for the study and thus possibly be assigned to the online course. Additionally, students majoring in non-math or science fields may find the course more challenging than students in STEM fields, in part because the former may not know as many faculty in the computer science department and thus might have

been less likely to ask for help when needed. At best, however, this is only a guess, and the fact that only 4 of the 140 students originally enrolled in CS 161 were declared computer science majors raises questions about whether a student's major is a significant predictor of his or her success in the course.

In any event, the collection of background data could have shed some light on the question of what student characteristics tend to be associated with better or worse performance in the different delivery formats. Knowing the ways in which students who were interested in taking an online version of CS 161 (i.e. the study participants) differed from those who were not interested in taking that online course (i.e. the nonparticipants) would also have been useful for understanding how generalizable the no-significant-difference finding (for the former group) might be to other settings.

Parsons-Pollard, Lacks & Grant (2008)

At an urban public university in the southeast with a total of enrollment of about 30,000 students, Parsons-Pollard, Lacks, and Grant examined a sample of 425 students in an online justice system survey course and 305 students who took a face-to-face version of that same course in the previous semester. In both the online course and the face-to-face courses, females outnumbered males, and of all the class years (freshmen, sophomore, juniors, and seniors), freshmen were the most well-represented. However, relative to the face-to-face course, the online course had a higher proportion of students who had not declared their majors or were majoring in a field outside of criminal justice.

In the online course, students were divided into five discussion groups in which they completed specific discussion forum assignments that involved analyses of relevant video clips. Online students also had several face-to-face meetings in which mini-lectures were given (rendering the term "hybrid" more appropriate than "online" for this group, though "online" is how the group is described in the article). Both the online and face-to-face groups used the same textbook and took very similar multiple-choice exams, and their test grades were weighted similarly. At the end of the semester, Parsons-Pollard and his colleagues compare the students' scores on three exams and their final course grades.

Although Parsons-Pollard and his colleagues do not explicitly identify withdrawal rates as an outcome of interest, they note that the withdrawal rates were "somewhat similar" for the two formats (2 percent in the online format and 8 percent in the face-to-face format). They, furthermore, find that while both the online and face-to-face group had acquired "sufficient knowledge" with respect to four of the five learning objectives within the course (though not the same four), the final course grades are significantly higher in the face-to-face group. However, online students did significantly better than the face-to-face group on the third exam. For the first and second exams, there is no significant difference between the two groups' performance.

Although Parsons-Pollard and his colleagues obtain a large sample, their design (comparing the performance of online students taking a course one semester to the performance of face-to-face students taking the same course a different semester) does not easily allow them to randomly assign students to one format or the other, leaving open the possibility of self-selection bias. In addition, while

the researchers note the differences between the students in the face-to-face course and the students in the online course, they do not appear to control for any of these differences in their analysis.

There are also some notable differences in the content and assignments given to the two groups. Whereas the face-to-face group used lectures and discussions only, the online students had access to websites on related topics, video interviews of professionals in the field, and discussion boards, the result of which, the authors note, might have been that the online course was more accommodating of “varying learning styles” than the face-to-face course. In addition, while the online students were assigned discussion forum exercises aimed at prompting critical thinking, the face-to-face students were not given similar assignments, precluding a comparison of the critical thinking skills of the two groups.

The researchers also report various types of technical problems that arose for the online students. Not only might have these impeded learning, but some of the problems might have made academic integrity violations harder to detect; the researchers note that at times it could be difficult to tell whether students who claimed they were “locked out” of the course were actually having technical problems, or whether they had been trying to cheat. Finally, the fact that the online students did not take tests in classrooms raises additional questions about the accuracy of the results.

Plumb & LaMeres (2011)

Plumb and LaMeres compare the performance of students in a face-to-face course titled Introduction to Logic Circuits at Montana State University, offered in fall 2010, to that of students in an online version of that course, offered in spring 2011. Introduction to Logic Circuits is a 16-week, sophomore-level course that typically enrolls 30 to 40 students a semester. While it is a requirement for students studying electrical and computer engineering, who collectively comprise about 90 percent of the course enrollments, it can be taken as an elective by students with other majors. The authors say little about the online course itself other than that it consisted of “modular materials developed summer 2010 for an online application.” The study states that the online class and the face-to-face class used the same materials, but that, for online students, these materials were accompanied by recordings of lectures, whereas the face-to-face students attended those lectures in person. While the student populations of the face-to-face and online courses were not significantly different from each other in terms of number of credits accrued, cumulative college GPAs, or SAT scores, the online students had, on average, significantly lower high school GPAs, were older, and were more likely to be non-degree or transfer students, relative to the students in the face-to-face class.

In comparing the performance of students in the two courses, Plumb and LaMeres find no significant difference between the two groups with respect to students’ final exam or final course grades. The researchers also examine student performance on eight modules, on each of which students were responsible for participating in one or more discussions, completing one or more homework assignments, and taking one quiz. For six of the eight modules, Plumb and LaMeres find no difference in the average scores of students in the face-to-face

and online sections. Of the two modules where the face-to-face and online students performed significantly differently, the online group did better on one module, and the face-to-face group did better on the other.

The results of this study, however, need to be interpreted with caution. Because the two formats were offered during different semesters, it is highly unlikely there was random assignment of students to one format or the other. More importantly, both the online and face-to-face sections had access to the same online materials; the only difference between the two formats seems to be whether the students viewed the lectures accompanying these materials in person or via recordings. Thus, there appears to be very little difference between the two formats; there is also little reason to expect that viewing a recording of a lecture and viewing a lecture live should result in significantly different learning outcomes.

Furthermore, although Plumb and LaMeres compare the student populations of the two groups, they do not appear to take any measures to control for the differences in these populations; they use only two sample *t*-tests in comparing the outcome measures of students in the two groups. Finally, while no sample sizes are explicitly given, even if all the students enrolled in the course participated in the study (about 40 students in each format), the sample size still would not be very large.

Poirier & Feldman (2004)

Poirier and Feldman randomly assigned 23 undergraduates at a large state university to a face-to-face or online version of an introductory psychology course. Of the 23 study volunteers, 12 were assigned to the online course, and the remainder to the face-to-face course, though two of the face-to-face students dropped the course and are consequently excluded from the analysis.

Students in the online course met in an online classroom several times per week for 15 weeks, accessing the course through the eCollege course delivery platform, which includes course materials such as text and audio files, reading assignments, web activities, short paper assignments, and practice exams, as well as communication tools. Twice a week, students participated in two online discussions, completed course activities, and received general feedback from the instructor regarding the course content. Students also communicated with each other and with the instructor through e-mail and threaded discussion. The students in the online section had no face-to-face lectures, and the only time they were required to come to campus for the course was when they needed to take exams (which they took at the same time as the students in the face-to-face section). By contrast, students in the face-to-face section had two meetings per week that lasted for 75 minutes each, and that consisted of a combination of lectures, videos, discussions, activities, in-class writing assignments, and exams. Face-to-face students were also asked to participate in three online discussions, using WebCT software to post comments, and to complete three short paper assignments. Students in both sections were given the same paper assignments and used the same textbook. The performance of the two sections on the four exams and three papers form the basis of the researchers' comparisons of the two formats.

Poirier and Feldman find that the online students answered a significantly higher percentage of exam questions correctly relative to the face-to-face students (an average of 79.6 percent correct for the former and an average of 71.9 percent correct for the latter), but that there is no significant difference in the students' grades on papers.

While this study's randomized assignment design makes it more rigorous than many of the other studies in this review, its findings should still be interpreted with caution because of the small sample size. Additionally, the fact that volunteers for the study had to agree to be randomly assigned to an online format or to a face-to-face format in order to participate might have caused the sample to be unrepresentative of all students enrolled in that course; indeed, the entire body of students enrolled in the course would also include nonparticipants, i.e. students who were not indifferent about the delivery mode in which they took introductory psychology. To their credit, Poirier and Feldman do compare the grades of participants and nonparticipants, and they report that there are no significant differences between the study participants and the nonparticipants, suggesting to them that the participants are "not atypical." However there is still the possibility that the participants and nonparticipants are different from each other with respect to unobservable characteristics, or with respect to observable measures of performance that are more fine-grained than overall course grades. Finally, the class size difference between the two formats—the traditional participants were taught in the context of a large-lecture format, whereas the size of the online section was very small—leaves open the possibility that any differences in outcomes may be caused more by the difference in class size than by the difference in delivery mode.

Reasons, Valdares & Slavkin (2005)

Reasons and his colleagues compare students' performance in purely face-to-face, purely online, and hybrid classes at the University of Southern Indiana, a public, four-year university that (at the time of the study) enrolled about 10,000 students. Two introductory-level undergraduate courses are involved in the study: EDUC 201: Introduction to Educational Psychology, and HP 211: Introduction to the Health Care Delivery System. HP 211 is a required course for students in the Bachelor of Science in Health Services program, while EDUC 201 is required for education majors. The researchers look at five traditional (purely face-to-face) sections, four online sections, and four hybrid sections of EDUC 201 and HP 211, offered over the course of six academic terms (which spanned three academic years). For EDUC 201, there were a total of 112 students enrolled in the traditional format, 22 students enrolled in the online format, and 40 students enrolled in the hybrid format. For HP 211, there were 96 traditional students, 54 online students, and 79 hybrid students. The students in the sample were mainly freshmen and sophomores. Two of the study researchers served as the instructors of the courses. However, "[t]o neutralize any bias that may have occurred, a third researcher participated fully in the study, and the retrospective analysis was designed to ensure that a formal and objective approach was undertaken."

In the purely face-to-face sections, students learned via lectures, discussions, and "hands-on" individual and group assignments. The instructor both distributed course materials in class and posted those materials on the Internet

for the students' convenience. The instructor administered exams in class; these exams consisted of multiple-choice questions from the "publisher's test bank" (where the "publisher" was presumably a textbook publisher).

In the hybrid sections of both EDUC 201 and HP 211, "students were expected to be involved approximately equally online and in face-to-face meetings." Students in the hybrid sections met seven or eight times over the course of the sixteen-week semester for lectures, presentations, and discussions. For group assignments, students completed some of the work face-to-face in class and some of the work online. Course materials and assignments were made available via the Internet. Discussions took place online; tests were administered online as well.

In the online sections, video-streamed lectures and lecture notes were posted on Blackboard, the online course management system; the instructor and students also used Blackboard to engage students in discussions through discussion boards and to administer tests. Online students' assignments generally consisted of "discussion forum reviews of material," group assignments, and individual projects which were submitted via e-mail. The students in the hybrid and online sections completed the same exams and assignments. There were no face-to-face meetings throughout the semester for the online sections.

The researchers compare the participation and final course grades of the students in the three different formats.²⁵ Final course grades were collected from university records; all final grades were based on a standard, university grading scale (with 90 to 100 percent amounting to an A, 80 to 89 to a B, 70 to 79 to a C, 60 to 69 percent to a D, and 59 and below to an F). Course participation was based on students' grades on their discussion board postings in the online and hybrid course sections, and based on students' participation grades in the face-to-face format. For both EDUC 201 and HP 211, the researchers noted that "a standard rubric that measured both the quality and quantity of the discussion board postings is used to support the reliability of course participation grades." The rubric was holistic and used a 1-to-5 scale (where 1 indicated "poor performance overall with little or no involvement" and 5 indicated "extensive participation with all assignments completed").

Before classes began, the instructors participated in a two-week intensive professional development program to prepare them to teach online education courses. The professional development covered basic instructional design, creation of online learning communities, student learning styles, organization and distribution of online learning materials, use of online discussion boards, e-mail management, copyright issues, integration of an effective assessment plan for online learning, and effective use of course management software.

The researchers find no significant differences among the different formats with respect to students' course participation. Comparisons of the final course grades of the students in the different formats show that students in the online

25 In addition, the researchers also compare the frequency with which students in the different formats interacted with the course website, measured by the number of times the site was accessed. However, this measure is concerned more with quantity of participation than with quality of participation and, thus, does not necessarily indicate the extent to which learning has taken place. Given that this literature review is concerned with comparisons of how well students learned the course material in the different formats, data regarding the frequency of students' participation in the different formats are not analyzed.

sections did significantly better than students in the traditional and hybrid formats, although there is no significant difference between the final course grades of students in the traditional sections and those of students in the hybrid sections. With respect to the “strength” of students’ participation levels, the researchers report that there are differences in participation among students, but no significant differences based on the format of the course. Reasons and his colleagues note that “a significant number of students in each of the sections of courses did not actively participate in activities and assignments,” which “disappointed” the instructors given that they “provided the infrastructure (grading rubric), detailed instructions on how to participate, and reminders that presupposed meaningful discussion activities.”

While the researchers are able to attain a respectable total sample size (403 students), the study has several drawbacks. For example, if students self-selected into their formats, as seems to have been the case since there is no mention of random assignment, a self-selection bias could confound the results. The researchers also do not collect any background data on the students (other than class year), preventing them or the reader from knowing whether the students in the different formats differ with respect to observable characteristics, including (though not limited to) past academic performance, income level, race/ethnicity, or age. Moreover, although final grades were based on a standard numbering system, nothing is said about whether final grades in the different formats were based on the same exams and assignments, aside from a mention of some standardization in grading with respect to the course participation outcome. Reasons and his colleagues also do not say anything about how heavily each of the components of the final course grade was weighted in determining the grade; for example, it could be that in the face-to-face section, exam performance was worth 60 percent of the final grade, compared with 40 percent in the hybrid section, and 30 percent in the online section. In addition, the tests and other assignments given to students in the different formats could also have differed substantially. In either or both of these situations, examining the final course grades alone would not necessarily allow for a fair comparison of the students in the different formats. Similarly, comparisons of participation based on students’ discussion board postings may be misleading, given that students in the traditional sections had fewer online assignments—course participation for traditional students was based not just on their activities on online discussion boards but also on their participation in class and on their attendance—and thus might understandably have devoted less attention and effort to their online discussion board postings than did students in the hybrid or online formats, who completed a greater share of their course work online.

Reasons and his colleagues identify a few other limitations to their study. One of these was the fact that some of the traditional EDUC 201 and HP 211 course sections were taught before the instructors attended the professional development program for online teaching and learning. Conceivably, this could prove to be a complicating factor to the extent that the workshop’s teachings might have made the instructors, who were teaching sections in all three formats, better overall teachers when they were teaching the online and hybrid sections after the professional development program, than they were when they were teaching some of the traditional sections before the professional development program.

Moreover, the researchers compare all participants in the face-to-face format, all participants in the hybrid format, and all participants in the online format. While this is not necessarily a flaw in the study, it would have been helpful if Reasons and his colleagues had disaggregated the results by course, to allow them to examine, separately, the relative effectiveness of the different formats for the education course and the relative effectiveness of the different formats for the health care course. The relative effectiveness of the different formats may be very different for the two courses, and those differences could be masked by aggregating the two courses' results.

It is likely (though not certain) that the same instructor taught all sections of EDUC 201, and a second instructor taught all sections of HP 211,²⁶ in which case the possibility of variation in instructor quality would not confound results that were disaggregated by course. However, under the current arrangement, in which the results are not disaggregated by course and in which the distributions of students among the three formats are different for EDUC 201 and HP 211, the instructor quality issue may prove problematic. To take an extreme example, suppose the instructor for EDUC 201 is excellent and the instructor for HP 211 is terrible. Of the 208 study participants who enrolled in the traditional format, 112 were enrolled in EDUC 201 and the remaining 96 in HP 211; of the 76 participants enrolled in the online format, 22 were enrolled in the EDUC 201 and the remaining 54 in HP 211; and of the 119 participants were enrolled in the hybrid format, 40 were enrolled in EDUC 201 and 79 were enrolled in HP 211. In this situation, the researchers would be comparing the traditional format—where 54 percent of the students in that format had the excellent instructor (that is, were enrolled in EDUC 201)—with the online format—where only 29 percent of the students had the excellent instructor—and with the hybrid format—where 34 percent of the students had the excellent instructor. One can see why such a comparison might not be conducive to isolating the effect of delivery format per se on student learning outcomes.

Rich & Dereshiwsky (2011)

Rich and Dereshiwsky conduct their study with about 100 students in four sections of an intermediate accounting course, titled Foundations of Accounting, at Central Connecticut State University. Two of the sections were face-to-face and met during the daytime; a total of 52 students participated in the study from these two sections. Another 30 participants came from a third section that also met face-to-face, but during the evening. The final section, a purely online section, had 19 students participating in the study; the online course used in this section was developed by one of the study authors, and the semester in which the study took place was the first time that online course was implemented. Of the students participating in the study, all but two were able to enroll in the delivery format of their first choice.

26 The article notes that two of its three authors were instructors in the study. Although it does not explicitly say that one of the two authors taught all of the EDUC 201 sections and the other instructor-author taught all of the HP 211 sections, this is probably a fair assumption, since one of the authors of the study is listed as a professor of education, and another is listed as a professor of "health services/administration."

The performance of the students in the different sections is compared based on their scores on nine “objective” online homework assignments, which were graded by the “publisher’s website” (presumably a textbook publisher, though this is not specified), and on their scores on a two- to three-paragraph essay on what it means to students to be a professional, which were graded by instructors using a rubric.

Of the four sections involved in the study, all except for the evening section were taught by the same instructor, and all study participants completed the same assignments. However, there were some substantial differences between the students enrolled in the different sections. Students enrolled in the online section were fairly similar to the students in the evening face-to-face section of the course: students in the online and evening face-to-face sections were more than twice as likely to be older than 22, much more likely to have full-time jobs, and more likely to be enrolled part-time, relative to the students in the daytime face-to-face sections. Compared with all three face-to-face sections, the online section tended to attract more students who were majoring in a field other than accounting.

Rich and Dereshiwsky find no significant difference in performance on homework assignments among students in the evening face-to-face section, daytime face-to-face sections, or online section, with one exception, which was that students in the evening section did significantly better than the daytime or online students on one topic covered in the course. For all topics covered in the homework assignments, there is no significant difference between the performance of the online students and that of the daytime face-to-face students. Students in the online sections did better than the students in the daytime or evening face-to-face sections on the essay on professional characteristics, although there is no indication of whether this difference is statistically significant.

There are several limitations to this study, the largest of which is that almost all the participants were allowed to self-select into the format of their choice, which resulted in different student populations enrolling in the different sections. In addition, the researchers make no effort to control for other student characteristics, such as demographic variables other than age (such as race, gender, and socioeconomic status) and prior academic performance indicators (such as GPA) that may be correlated with course outcomes. Moreover, without the random assignment of participants into the different formats, the possibility of further, unobservable differences affecting the outcomes cannot be eliminated. The study also would have been improved had the researchers controlled for instructor effects, even in the presence of objective or standardized ways of grading assignments, and included among the outcome measures scores from one or more proctored tests, for which the researchers would be better able to guard against the possibility that the participants were submitting work that was not their own.

Riffell & Sibley (2005)

In this study, Riffell and Sibley compare the post-test scores and performance on an active-learning activity of 55 students who were enrolled in a hybrid introductory biology course, with those of 74 students who were enrolled in a face-to-face version of that same course. The participants were all resident, non-science majors at Michigan State University who were drawn from the 102 students taking Applications of Environmental Biology in a traditional format, and from the 85

students taking the course in the hybrid format during the semester in which the study took place. Students were allowed to enroll in the section of their choice, but they did not know about the study at the time they were registering.

The students taking the hybrid version of the course had one 50-minute face-to-face session each week, described in the study as an “active lecture.” During an “active lecture,” a non-tenure-track instructor gave brief lectures at the beginning and the end of the sessions, and, in between, students worked together in groups to complete “active-learning” exercises. In addition, the students in the hybrid course completed bi-weekly homework problem sets, each containing about 50 questions, using a web-based, open source platform called LON-CAPA (<http://www.lon-capa.org/>). LON-CAPA provided the students with individual questions (which they were allowed multiple attempts at answering correctly), with pre-programmed hints and instant feedback; LON-CAPA also graded all the online assignments. By contrast, students in the traditional, face-to-face format of the course spent three hours each week in lecture: they participated in one “active-learning” session each week, plus two additional hours of more passive lecturing, in which instructors covered the same material that the hybrid students learned through their homework assignments in LON-CAPA, but did not “directly query” students.

Both the face-to-face course and the hybrid course covered the same subject matter, and were taught by the same instructor. The two classes were also similar to each other in terms of the gender composition, the proportion of students who had previous experience with online courses, the proportion of students who were full-time rather than part-time students, and the proportion of students who had previously taken an integrative studies biology course at Michigan State. However, in other respects the populations differed: the face-to-face group contained significantly more freshmen and significantly fewer commuter students.

In their analysis, the researchers control for gender, performance on a pre-course assessment, attendance at the active-learning lectures and at the more passive lectures, previous experience with online courses, commuter student status, class year, and active-learning activity performance. They also include as controls the interaction of delivery format with gender, the interaction of delivery format with previous experience with online courses, and the interaction of delivery format with class year.

Riffell and Sibley find that the students in the hybrid group performed better on the post-course assessment than did students in the face-to-face group; while both upperclassmen and freshmen did better in the hybrid class than in the face-to-face class, the difference is significant only for upperclassmen. In addition, the researchers find that, among the questions on the post-test assessment that pertained to material covered by the active-learning sessions, hybrid students did better than face-to-face students, suggesting that active-learning exercises were more effective when they were coupled with online activities than when they were coupled with passive lectures. None of the interactions of format with covariates is significant for the post-test questions related to content covered in the active-learning sessions.

While Riffell and Sibley are more diligent about controlling for differences in the student populations in the face-to-face and hybrid classes than the authors of a number of the other studies in this review, there are other controls which might have helped to isolate the effect of the delivery format that they could have included but did not, such as race or ethnicity, age, and GPA. In addition, allowing students to self-select into the format of their choice, even if the students did not know about the study when they registered for the class, leaves open the possibility of unobservable differences in the student populations confounding interpretation of the results.

Samman et al. (2012)

In this study, Samman and his colleagues compare the outcomes of freshmen in online and face-to-face versions of three different introductory courses at a public four-year institution over the course of five semesters. (All four authors of the article describing the study are affiliated with Southern University at New Orleans, suggesting this might have been the site of the research.) The introductory courses involved in the study are BIOL 105: Introduction to Biology, ENG 111: English Composition, and COMM 210: Fundamentals of Public Speaking. The researchers examine the pass rates and mean distributions of grades for these classes for the following semesters: spring 2008, fall 2008, spring 2009, fall 2009, and spring 2010.

In the spring and fall semesters of 2008, there were a total of 367 students in BIO 105 enrolled in the face-to-face format, and 91 BIO 105 students enrolled in the online format. (The researchers do not provide the breakdown of students between the spring and fall semesters.) In 2008, there were a total of 202 students enrolled in the face-to-face format of COMM 210 and 59 students were enrolled in the online format of that class. For ENGL 111, 509 students were enrolled in the face-to-face version and 136 students enrolled in the online version in 2008.

In 2009 there were 404 students in the face-to-face version of BIO 105 and 104 participants in the online version of that course; 180 students in the face-to-face version of COMM 210 and 58 students in the online version; and 456 students in the face-to-face version of ENGL 111 and 61 students in the online version.

In 2010, there were 404 students in the face-to-face version of BIO 105, and 53 students in the online version of BIO 105; 181 students in the face-to-face version of COMM 210 and 43 students in the online version of COMM 210; and 506 students in the face-to-face version of ENGL 111 and 93 students in the online version of ENGL 111.

The institution's Information Technology Center provided researchers with the participants' grades in the courses involved in the study for the relevant semesters. The researchers compare both the grades and the passing rates of students in the two formats for all three courses (where A, B, C, and D are passing grades and F is a failing grade). Comparisons of the mean distributions of grades over the five-semester period of the study using *t*-tests show that there is no significant difference between the two formats for COMM 210 (the *t*-score for which is 1.35). However, for ENGL 111, the face-to-face students did significantly better than the online students; the difference has a *t*-score of

2.50 and, with a p -value of 0.0125, is almost significant at the 1 percent level. In addition, in BIO 105, students in the face-to-face courses outperformed their online counterparts; the difference is significant at the 0.01 percent level, with a t -score of 3.93.

In addition to comparing the grades of the students in the two formats, the researchers also compare the pass rates in each course for 2008, 2009, and 2010. On the whole, the face-to-face versions have higher pass rates than their online counterparts, although there is some variation across the three courses. For BIO 105, the face-to-face format's pass rate held steady between 2008 and 2010, whereas the pass rate of students in the online format increased during that time period, starting lower than the pass rate of the students in the face-to-face format but ultimately surpassing it. The pass rate for the face-to-face version of BIO 105 was 54 percent in 2008, 53 percent in 2009, and 55 percent in 2010, compared with, for the online version, 34 percent in 2008, 47 percent in 2009, and 64 percent in 2010. With respect to COMM 210, the pass rate changed little in either format over the period of the study. The face-to-face students passed at a rate of 71 percent in 2008, 75 percent in 2009, and 74 percent in 2010, compared with slightly lower pass rates for the online students during each year of that period: 66 percent in 2008, 72 percent in 2009, and 71 percent in 2010. Finally, in ENGL 211, like in BIO 105, the face-to-face students' pass rate did not change substantially over the period of the study, whereas the pass rate of the students in the online format increased over that period (although the highest pass rate for the online students is still lower than the lowest pass rate for the face-to-face students). In ENGL 211, 44 percent of face-to-face students passed in 2008, 43 percent passed in 2009, and 48 percent passed in 2010, compared with pass rates for online students of 29 percent in 2008, 36 percent in 2009 and 40 percent in 2010.

While this study involves fairly large sample sizes, the researchers' methodology leaves much to be desired. For example, students appear to have been allowed to self-select into the format of their choice, and the researchers mention no attempts to control for the possibility that the students who were enrolled in the different formats are different in ways that might affect their grades or pass rates (both with respect to relatively easily observable variables, such as past academic performance, and with respect to less easily observable variables, such as work ethic and motivation). Thus, the results are likely to be distorted by a selection bias. The researchers also do not appear to have controlled for differences in instructor quality or to have standardized the content or assessments in the face-to-face and online versions of the course to make the courses as comparable as possible. For all the reader knows, the face-to-face and online versions of the courses could have covered substantially different content and involved substantially different assignments and assessments, which were graded according to different standards and were taught by instructors with very different skill levels. In such a scenario, any differences in outcomes between the two formats could be attributable to differences in content, assessment, grading, and/or instructor quality to a greater extent than to differences in the delivery format per se.

Finally, comparing final course grades is not necessarily helpful without knowing the assignments or other factors (like class participation) on which those final course grades were based, or the grading standards or rubrics (if any) used by the instructors. For example, an instructor teaching in one format may assign

her students only essays throughout the semester and grade these essays very leniently, whereas another instructor, teaching in a different format, may be a much stricter grader and give her students much more challenging assignments and exams. In such a situation, merely observing that the grades of the students taught by the first instructor tend to be higher than those of the students taught by the second instructor does not constitute grounds for concluding that the format in which the first instructor taught is a more effective format than that in which the second instructor taught. Comparisons of only the final course grades and the pass rates (which are, in turn, based on the final course grades) of students taking a given course in two different formats would not necessarily allow researchers to determine the relative effectiveness of those formats.

Schwartz (2012)

In an observational study examining data collected in previous years, Schwartz compares the performance of students in 62 face-to-face sections with that of students in 70 online sections of various accounting classes. The participants of this study are students who were enrolled in bachelor's accountancy programs at the National University in San Diego, which caters mainly to adult learners (the average age of students at the university is 32). The online sections in this study involved asynchronous threaded discussions along with at least 90 minutes of live chat sessions per week. In 2004, a voice-over IP (VoIP)-based system was adopted to enable voice interaction between students and the instructor (through the use of headsets with microphones), as well as visual presentations using whiteboards, PowerPoint, spreadsheets, and other media. Instructors were encouraged to offer two live chat sessions per week, which were recorded and which the students could play back at their convenience. In practice, these sessions often exceeded two hours. The face-to-face sections, by contrast, generally involved 40 to 45 hours of face-to-face time per course. The researchers say there are "no significant demographic differences between students in online classes and their counterparts in onsite courses," and that, within a course, all sections used the same syllabus and the same textbook, and were based on the same learning objectives.

In this study, Schwartz compares students' performance on standardized tests as well as students' course grades. The standardized tests used in this study, the Standard Learning Outcome of Achievement Test (SLOAT), have been used in the National University's business school since 2003. Department faculty prepare the SLOAT tests and send them to instructors a few weeks before the end of the course. In this study, SLOAT tests were administered in the online and face-to-face sections of many (though not all) of the courses in the accounting program, including federal taxation, cost/managerial, auditing, and intermediate-level financial accounting courses.

In total, Schwartz analyzes scores from 189 tests taken by students in 19 face-to-face sections, taught by nine different instructors, and scores from 372 tests taken by students in 19 online sections, taught by seven different instructors. (Schwartz notes that, because many students took tests in more than one of the courses involved in the study, 189 and 372 do not represent the number of face-to-face participants and online participants, respectively, whose data are analyzed in the study. Rather, "since two to three onsite sequences and two online sequences of B.S. in Accountancy courses [the degree program

of which all the courses involved in the study are part] are offered each year, it is estimated that over the two-year period for which SLOAT [standardized test] scores were analyzed, the number of different students who took these tests was approximately one-quarter the number of tests administered.”)

Using *t*-tests for his analysis, Schwartz finds that, overall, the mean course grades and standardized test scores are significantly lower for students in the online sections than for students in the face-to-face sections. The online students earned a mean SLOAT score of 2.80 and a mean course grade of 2.86 (both graded on a scale of 0 to 4.0), and the face-to-face students earned a mean SLOAT score of 2.93 and a mean course grade of 3.00; both differences are significant at the 5 percent level. However, when Schwartz disaggregates the results into four major subject areas—results for financial accounting classes, for tax accounting classes, for cost/managerial accounting classes, and for auditing classes—he finds few significant differences in the performance of the students in the two formats in most individual subject areas. Specifically, there are no significant differences in course grades or test scores for financial accounting courses, auditing courses, or cost/managerial courses. (In the financial accounting courses, the students in the online sections earned a mean SLOAT score of 3.05 and a mean course grade of 2.75, compared with a mean SLOAT score of 2.99 and a mean grade of 2.82 for students in the face-to-face sections. In the cost/managerial courses, students in the online sections earned a mean SLOAT score of 2.85 and a mean course grade of 2.88, compared with a mean SLOAT score of 2.93 and a mean grade of 2.88 for students in the face-to-face sections. Finally, in the auditing classes, students in the online sections earned a mean SLOAT score of 2.92 and a mean course grade of 2.98, compared with a mean SLOAT score of 3.02 and a mean grade of 3.12 for students in the face-to-face sections. None of these differences is significant at the 10 percent level.)

However, for “income tax” courses (which the author seems to use synonymously with “tax accounting” courses), the face-to-face group did significantly better than the online group with respect to both SLOAT scores and course grades. (The face-to-face students achieved a mean standardized SLOAT score of 2.46 and a mean course grade of 3.25, compared with a mean SLOAT score of 2.01 and a mean course grade of 2.78 for students in the online format; the difference in SLOAT scores is significant at the 5 percent level, and the difference in course grades is significant at the 1 percent level.) When Schwartz looks at the results excluding income tax courses, he finds that the performance of the students in the online and face-to-face formats is not significantly different. (When data from income tax courses are excluded, face-to-face students had a mean SLOAT score of 2.98 and a mean course grade of 2.96, compared with a mean SLOAT score of 2.99 and a mean course grade of 2.87 for online students; neither difference is significant at the 10 percent level.)

Schwartz’s comparisons of online learning and face-to-face learning should not necessarily be viewed as informative. One reason is the limited amount of standardization between the two formats: instructors of a given course were free to use their own teaching strategies, assignments, exams, and grading scales, thereby raising the question of whether these differences may serve as confounding factors. In addition, students in the face-to-face classes took closed-book tests, whereas students in the online classes took open-book tests (although

Schwartz says the online classes' exams had "a time limit that impedes students' ability to look up answers in the book" and consisted of questions that were "designed to preclude the finding of answers in the textbook or lecture notes").

In addition, the fact that this was a non-randomized study that allowed students to self-select into the format of their choice raises the question of selection bias. Schwartz acknowledges as much, noting one factor that may have made some students more inclined to opt for the online format rather than the face-to-face format, and thus suggesting one possible difference between the groups of students who self-selected into the different formats that may be relevant to the students' course performance: "Many students opt for online classes because they perceive (mistakenly) that online courses are less demanding of their time, which, for the working adults who comprise the majority of this university's student body, is in short supply." Another confounding factor in this study is a difference in class size: the average class size of the online sections was more than twice as large as that of face-to-face sections. Finally, Schwartz's mention of "less-than-thorough training and monitoring of [online] instructors on adjustments to teaching strategies and on ways to make effective use of the online functionality" suggests there may have also been differences between the two formats with respect to the teaching quality. In light of these concerns, the results of this study certainly cannot be called conclusive.

Signor (2003)

Signor compares the raw final exam scores of students in traditional and hybrid versions of a database concepts and modeling course over a four-year period at Swinburne University of Technology in Australia. The course involved, LAI 210: Database Concepts and Modeling, is a second-year course whose prerequisites include two information technology classes in which students learn basic computer literacy concepts. The courses examined in the first two years of the four-year study period (1999 and 2000) were taught in a traditional face-to-face format and constitute the control group; the courses taught during the second two years (2001 and 2002) were taught in a hybrid (part online, part face-to-face) format and constitute the treatment group. The face-to-face classes consisted of lectures, which were delivered in a theater, and "tutorials," which took place in classrooms; students were responsible for taking their own notes and for printing computer-based course materials. In the hybrid format, students had access, via the Internet, to virtual lectures—that is, to PowerPoint presentations that contained "text, image and audio capabilities"—and also attended the face-to-face tutorials.

In the article describing her study, Signor notes that there are "several extraneous variables that threaten the validity of studies of this nature" and that those "were identified and where possible, minimized." One such variable is possible changes in the course content over the time period of the study. However, Signor says, because LAI 210 focuses on historical data-based concepts and traditional modeling tools, its "content has remained relatively static and the exam has not changed significantly" during the study period. In addition, the length of the tutorials in the hybrid format was increased so that the hybrid students in the study had as much face-to-face time with their instructor each week as did the students in the purely face-to-face format; the purpose of this was to prevent

the exam performance of students in the hybrid format from being “adversely affect[ed]” (relative to that of the students in the face-to-face format) by their having less in-person contact with their instructor.

When Signor compares the final raw exam scores of students in the face-to-face and hybrid LAI 210 classes over the four-year study period using an independent *t*-test, she finds no significant differences between the scores of the students in the two formats. (The students in the hybrid format earned a mean score of 59.32 on the final exam, with a standard deviation of 12.35; the students in the face-to-face format earned a mean score of 60.37, with a standard deviation of 12.05.)

The results of this study do not offer very compelling evidence about the relative effectiveness of the two delivery formats. For instance, Signor does not provide any sample sizes, preventing the reader from knowing whether there were 20 students enrolled in LAI 210 in each of the four years involved in the study, or 200 students, or whether the number of students enrolled differed substantially from one year to the next; in any event, the sample size would be very useful to know in interpreting the study results. All participants were also allowed to take LAI 210 in the format of their choice, allowing for self-selection bias to confound comparisons of the students in the different formats. Signor also does not appear to have collected any background data on the participants and thus is unable to control in her analysis for background variables that might have affected students’ exam performance.

Signor mentions that the content of the course and the final exam did not change substantially during the four-year study period; that the students in both face-to-face and hybrid formats had the same amount of face-to-face contact with the instructor each week; and that the course’s prerequisites ensured that the students enrolled would have similar baseline levels of computer literacy. However, she does not indicate whether the same instructor taught the course over the four years of the study period, or cite any attempts to control for possible differences in instructor quality. She also does not say anything about standardization (across the two formats) of variables like section size, course assignments, reading materials, or other components of the course. Thus, given the existence of self-selection bias and possibly of other confounding variables, and the author’s failure to provide sample sizes, one should not place very much weight on the results of this study.

Summers, Waigandt & Whittaker (2005)

Summers, Waigandt, and Whittaker compare the academic performance of students in online and face-to-face versions of an undergraduate introductory statistics course. Thirty-eight students, enrolled in the School of Nursing at a “large midwestern university,” participated in the study. (The authors subsequently provide a brief overview of distance education at the University of Missouri, suggesting that this might have been the university involved in the study.) Twenty-one of the thirty-eight students chose to take the course in a face-to-face format, and the remaining seventeen opted to enroll in the online format. Most of the students in the sample were “upper-division” students for whom the statistics class was a required course.

The online version of the class, prepared over the course of three months, was created by an instructor who had traditionally taught the course face-to-face. According to the authors of the study, this instructor “had had so much past success with the delivery of the material in a face-to-face environment” that “he literally translated all of his face-to-face lectures, activities, and homework assignments to the online format.” The online course was delivered via the delivery system WebCT and was “structured to look like a website”; the interface had a row of buttons on one side, which students could click on to access different components of the course, such as the syllabus, assignments, and supplementary materials. Students in the online format typically received reading and written assignments each week, which they completed and then submitted via e-mail. E-mail was the sole medium through which the instructor of the online format interacted with his students and through which students in the online class interacted with each other; students generally responded to questions posed by the instructor in threaded e-mail discussions. Both the online and face-to-face versions of the course were taught by the same instructor and involved the same content and examinations.

The researchers compare the performance of the students in the two formats using a cumulative score that is based on three “class exams” and on a final exam. Both the online and face-to-face students’ exams consisted of a total of 175 items and tested students on the same content. Students in the face-to-face class took these exams during their class time, whereas students in the online format took the exams at a location near them, under the supervision of an approved proctor, in connection with the university’s Direct Distance Learning Program. Students in both formats had the same amount of time to complete each exam.

At the beginning of the semester, students in both formats were given a pre-test to provide researchers with a baseline indication of their skills; the results of the pre-test showed no significant difference between the two formats. (The *t*-test used to compare students’ performance on the baseline assessment and resulted in a *t*-score of 1.25.) At the end of the 15-week course, when researchers compare the cumulative test scores of the students in the two formats using a *t*-test, they also find no difference between the two formats with respect to students’ statistics knowledge (the *t*-score for this test was 1.42).

The researchers deserve credit for their efforts to standardize the content and assessments in the two formats, for administering a pre-test to measure the baseline statistics skills of the participants, and for having the same instructor teach both formats to prevent differences in instructor quality from confounding a comparison of the two formats. However, the study still has a fair number of deficiencies. Summers, Waingandt, and Whittaker acknowledge that the research was “limited by using a small convenience sample of students.” Not only is the sample size small, but the researchers also allowed students to self-select into the format in which they took statistics. While the students in the two formats do not differ significantly with respect to their performance on the pre-test, the use of this pre-test score alone is unlikely to be sufficient to demonstrate that the students who self-selected into the two formats are comparable in as far as their relevant academic ability is concerned. There may also be other background variables with respect to which the students in the two formats differ. For example, the students in one format may come from substantially wealthier families than the

students in the other format, in which case the researchers may not have failed to find a significant difference in pre-test performance had they controlled for family income. If the two groups of students were to score significantly differently on the pre-test once family income is controlled for, they may not be as similar with respect to academic ability as the researchers suggest. In that case, the researchers would be well-advised to control for differences in baseline abilities in examining the performance of students at the end of the course; these controls would allow the researchers to better isolate the effect of the course format on students' exam performance, from the effect of students' background abilities on their exam performance. In any event, in light of the fact that the research involves a "small convenience sample of students" who were allowed to self-select into the format of their choice, and the fact that Summers and his colleagues do not attempt to collect data on, or control for, background variables with respect to which the students in the two formats may differ, the researchers' finding of no significant difference in students' outcomes is not very compelling.

Urtel (2008)

In this study, which took place at a local Midwestern urban university, Urtel compares face-to-face and online instruction in an unidentified course. The face-to-face version of the course involved in the study contained 116 students, and the online version contained 269 students, for a total sample size of 385 students, all of whom were undergraduates. This course historically has been offered as a purely face-to-face course, although a few years before the study was conducted, the course was converted into an online format so that both face-to-face and online versions of the course could be offered. The course has no prerequisites.

In the face-to-face version of the course involved in this study, students met two times a week for 75 minutes each time, over a 15-week period. The face-to-face time was used "to accomplish the learning objectives for the reading and lecture content of each session." In the online format, students accessed the lecture materials either through links to videotaped lectures that were posted online, or by watching the lectures on local cable television. Online students also participated in discussion forums and chat rooms. Relative to students in the face-to-face format, students in the online format were "equally prompted for student interaction and discussion so that, conceptually, learning objectives from the reading and lecture content could also be discussed." Both the face-to-face and online versions were taught by the same instructor and used the same course content, textbook, reading and lecture schedule, and "overall performance assessments." In both the face-to-face and online versions of the course, "student interaction and discussion was [sic] emphasized."

Urtel analyzes data from final grade reports and from electronic or physical grade books. His outcomes of interest are students' final course grades and DFW rates (rates of students earning a D, an F, or a W—for withdraw—in a class). In addition to performance data, he also collected demographic data (including student gender, ethnicity, class level, and age) and other background data (on variables such as GPA).

When he compares the face-to-face and online formats using factorial analysis of variance, Urtel finds a significant difference between the course grades of the students in the two formats, with students in the face-to-face format earning

grades that are significantly higher at the 1 percent level. (The face-to-face group earned an average grade of 3.16 on a 4.00 scale, whereas the online group earned an average of a 2.28 on the same 4.00 scale.) Disaggregating the data by class level shows that the difference is most extreme for freshmen. (While the exact figures for freshmen are not provided, from a graph the average grades look to be about 1.8 for online students and about 3.15 for face-to-face students.)

When comparing the DFW rates of the students in the different formats, Urtel finds that students in the online format have significantly higher DFW rates than the students in the face-to-face format (the former group has an overall DFW rate of 40 percent, compared with a 21 percent DFW rate for the latter group). There is no significant difference between the different formats with respect to the age of the students who withdrew from the course, and in each format, the DFW rates are higher for freshmen (who constitute about 20 percent of the total course enrollment in each format) than for the class as a whole. However, among freshmen, the DFW rate is significantly higher in the online format than in the face-to-face format. (The freshmen in the online group have a DFW rate of about 65 percent, compared with about 35 percent for the freshmen in the face-to-face group.)

With respect to the effects of ethnicity, gender, and age, Urtel finds that, among white students, those who took the course in the face-to-face format did significantly better than did those who took the course in the online format. White students who took the course in the face-to-face format earned an average grade of about 3.6, compared with an average grade of slightly less than 3.0 for white students who took the course in the online format; with a *p*-value of 0.0132, this difference is almost significant at the 1 percent level. However, there are no significant differences (even at the 10 percent level) by format for black or Hispanic students (with the former group earning an average grade of about 2.25 in the face-to-face format and 2.0 in the online format, and the latter group earning an average grade of about 3.35 in the face-to-face format and 3.1 in the online format). When looking at gender subgroups, Urtel also finds that female students in the face-to-face format did significantly better than female students in the online format (with females in the face-to-face format earning an average grade of about 3.7, and females in the online format earning an average grade of about 2.9; the *p*-value for this difference is 0.014). The difference between the two formats is not significant for males. (Males earned an average grade of about 2.4 in the online format and an average grade of about 3.4 in the face-to-face format; the *p*-value for this difference is 0.106.)

Compared with a number of the other studies in this review, this study has a fairly large sample size and supplements its comparison of students' performance with collection of a decent amount of background data, which are incorporated into Urtel's analysis. In addition, Urtel deserves praise for his study's standardization of many aspects of the course to ensure that his comparisons isolate the effects of the delivery format to the greatest extent possible, by holding constant the course content, textbook, reading and lecture schedule, and overall assessments, and by having the students in both formats taught by the same instructor. This being said, even if the assessments given to the students in the two formats were the same, it is still unclear whether, in the different formats, the instructor gave the same weights to those assessments in determining

students' final grades. (Although the instructor may have standardized this component of the course as well, it is not specified in the article; it would have been helpful to know whether this was the case.)

The biggest deficiency in the study, however, is the fact that students do not appear to have been randomly assigned to one format or the other. The differences between the students in the different formats that are evident from the demographic data Urtel presents suggest that there may be further differences related to students' academic performance. (For example, the distribution of students among the different class levels differs noticeably—unfortunately, Urtel does not indicate whether the differences are significant. The proportion of students who are seniors is much larger, and the proportion of students who are sophomores is much smaller, in the online format than in the face-to-face format.) Although Urtel describes the demographic differences between the two formats, he does not attempt to control for them, making it very difficult (if not impossible) to isolate the effects of the delivery format on students' academic performance, from the effects of pre-existing differences in the students who self-selected into the different formats. Finally (although this is not necessarily a flaw in the study), it would also have been helpful, in interpreting the results, to know the subject of the course involved in the project, as the effects of the online format could differ by discipline.

Wagner, Garippo & Lovaas (2011)

Wagner, Garippo, and Lovaas conducted a longitudinal observational study of students who were enrolled in sections of an undergraduate introductory course in the use of business application software offered between fall 2001 and spring 2010. Their sample includes 435 students in 19 face-to-face sections and 171 students in 11 online sections, excluding 18 students who withdrew from the study. Fifty-two percent of the participants are female. The same instructor taught both the face-to-face and online formats, using the same syllabus, giving the same assignments and course exams, and grading according to the same standards. An independent samples *t*-test shows no significant difference between the face-to-face students and the online students with respect to final grade percentage. The researchers also find a gender main effect (significant at the 5 percent level) when they look at the data with gender and course delivery factored together, with each gender earning lower average grades in the online format than in the traditional format (an average final grade of 84.2 percent versus 87.7 percent for males, and 88.1 percent versus 89.7 percent for females).

However, this study has a number of limitations. The researchers allowed students to self-select into the format they wanted, and they do not examine the effects of the delivery format on any other subgroups other than males and females, nor do they attempt to identify or control for other differences in the student populations of the face-to-face and online sections. There also do not appear to be any attempts during the study to make sure there were no academic integrity violations, particularly by online students who were not “directly supervised” while taking exams or completing assignments (on which students' final grade percentages were based).

Weber & Lennon (2007)

Weber and Lennon conduct two similar studies at two private southeastern universities, both involving sections of a junior-level course titled Principles of Marketing. In Study 1, the researchers compare the performance of 31 face-to-face students to that of 17 online students. In both the face-to-face and online groups, females outnumber males but, relative to students in the face-to-face section, students in the online section have significantly lower GPAs, significantly greater experience with online course components, and greater perceived knowledge of web-based courses. In Study 2, Weber and Lennon compare the performance of 35 face-to-face students to that of 34 online students. As is true in Study 1 sections, females outnumber males in both the face-to-face and online sections in Study 2; in addition, juniors make up more than half of both groups in Study 2. As in Study 1, the students in the online sections in Study 2 have slightly lower GPAs, slightly greater experience with online course components, and a higher perceived knowledge of web-based courses. In general, the students in Study 2 have more experience with web-based courses than the students in Study 1.

Students in the web-based sections of the course used “a modular format which presented information and incorporated learning activities” each week, and had access to all lectures and course materials via websites. Projects assigned to the web-based sections generally focused on “some type of business entity with an online perspective.” The learning activities and information presented to these sections were said to be comparable to those presented to face-to-face sections, and the instructor, textbook, exams, assignments, projects, and material coverage were the same across all sections.

At the end of the semester, the researchers compare the two groups’ scores on the final exam (which consisted of 50 multiple-choice questions and three short-answer essays) and on a semester project, and students’ overall course grades (which were based on their grades on exams and on the semester project). In addition, Weber and Lennon collected information via questionnaires at the beginning of the semester about student GPA, gender, class year, “web-based experiences” in classroom settings, and perceived knowledge about the Internet. The researchers then look at the extent to which these predictor variables (as well as course format) are correlated with outcomes.

Weber and Lennon find that, relative to the face-to-face sections, the online sections have a higher drop-out rate, although they do not indicate whether the difference is significant. Online students also did significantly worse than the face-to-face sections in both Studies 1 and 2 with respect to the project grade, final exam grade, and the final course grade. However, when the researchers run regressions which include controls for predictor variables, they find that course format is not a significant predictor of “learning achievement.”

Although the authors attempted to reduce self-selection bias by not announcing that one section would be online until the first week of class, there was no random assignment. Students were allowed to switch sections upon finding out whether their section would be web-based if they had a “strong objection” to this format. The fact that no students acted on this option does not negate the fact that the option was available, and the fact that the students elected to remain in their

assigned sections still reveals a selection bias. Although Weber and Lennon collected information about and controlled for, some predictor variables, they do not appear to have collected information about other predictor variables that may be correlated with student performance, such as socioeconomic status, race or ethnicity, and number of hours worked per week for pay. Finally, the sample sizes, while not extremely small, are still not large enough for Weber and Lennon's studies, by themselves, to offer conclusive evidence about the relative effectiveness of different delivery formats.

Wilson & Allen (2011)

Wilson and Allen's research involves 101 undergraduate students at a historically black college who were enrolled in at least one of two management courses. One of these courses, Management 311: Principles of Management, covers the "theory and process of management principles and basic organization behavior"; the other course, Management 420: Organizational Behavior, covers "human behavior at the individual, group, and organizational level." In this study, both courses were offered in a face-to-face format as well as an online format, with students enrolling in the format of their choice. A total of 58 students enrolled in the online courses; a total of 43 students enrolled in the face-to-face courses. Wilson and Allen collected student course grades and several variables relating to students' backgrounds, such as gender, race (where students were categorized as either black or non-black), high school grades, SAT scores, whether the student initially enrolled at the institution as a first-time student or as a transfer student, age, cumulative college GPA, cumulative credit hours, and whether the student was "lower- or upper-division" (presumably indicating whether the student was in his freshman or sophomore year of college, or whether he was in his junior or senior year).

Of the 58 students who enrolled in the online format, 14 withdrew from their courses and 44 completed their courses; of the 43 students who enrolled in the face-to-face format, 6 withdrew from their courses and 37 completed their courses. Wilson and Allen perform a 2-by-2 analysis of variance (ANOVA) using two variables: completion status (broadly defined in terms of course completers versus course withdrawers) and delivery mode (online versus face-to-face). Among the students who completed their courses and thus earned course grades, there is no significant difference in the course grades of the students in the two formats: students in the online format earned an average course grade of 2.09 (presumably on a 4.00 scale), and students in the face-to-face format earned an average course grade of 2.14. (This difference is not significant at the 5 percent level, which is the researchers' threshold for significance).

This study, unfortunately, has a number of limitations. First, the sample size is not very large (although it is not as small as the sample sizes of many other studies). Although there are 101 students in total across both formats and both management courses, 20 of these students (14 online students and 6 face-to-face students) withdrew from their respective courses. The researchers have course grades for the 81 students who completed their courses—44 online students and 37 face-to-face students. However, the researchers do not specify how many of the 37 face-to-face students, or how many of the 44 online students, are in the Management 311 course and how many are in the Management 420 course. It

could be, for example, that an overwhelming majority of the online students are in Management 311, but the overwhelming majority of face-to-face students are in Management 420. In that case, comparing a sample of online students who are disproportionately enrolled in one course with a sample of face-to-face students who are disproportionately enrolled in a different course would not result in a fair assessment of the relative effectiveness of the two delivery formats.

Moreover, Wilson and Allen do not indicate any attempt to standardize, between the face-to-face and online sections of each management course, the course content or the exams, leaving open the possibility that the grading standards and assignments used in the two formats are not comparable. It is also unclear whether the instructor was the same in the face-to-face and online sections of each course; if not, and if the researchers make no attempt to control for differences in instructor quality, these differences may confound comparisons of the different formats.

Finally, while the researchers collected data on, and include in their analysis, a fairly extensive array of demographic variables, they were unable to collect data on other student characteristics that could be relevant to student performance. Examples of observable variables omitted from the analysis include income level, parental education, part-time versus full-time enrollment status, and hours worked for pay (if any), during the semester in which the study took place. Moreover, even the background variables about which the researchers did collect data are not included as covariates in regressions of course grade on format. Instead, it appears that Wilson and Allen regress course grade on background variables for the online students, and course grade on background variables for the face-to-face students, separately. Running separate regressions for the face-to-face and online students shows the relationship between course grade and the background covariates, but it does not show the relationship between course grade and delivery format, which is the relationship of interest for the purposes of this literature review. This being said, even if the researchers did perform regression analysis using a more comprehensive set of background variables as covariates, they could not be completely sure that the students who self-selected each format did not differ from each other in ways that are unobservable but still related to academic performance—a problem that could have been solved had the researchers randomly assigned students to one format or the other.

Xu & Jaggars (2011)

Xu and Jaggars analyze data from 18,896 students who lived in Washington State and who, in fall 2004, entered one of the colleges in the Washington State Community and Technical College System with the intent to transfer to a four-year college. The researchers follow the students from fall 2004 through summer 2009, collecting information about 126,317 course enrollments over this five-year period. None of these enrollments is in a remedial course, and none of the enrollments is in a course for which only one delivery format of that course was available at the student's campus. Approximately 22 percent of the enrollments examined in this study are for "online" courses, which, in this study, means courses in which at least 51 percent of the instruction and student-teacher interaction took place online. The researchers also collected an assortment of background data on students. Based on these data, they determine that, relative

to the students who took courses in the traditional format, the students who took courses online were more likely to be older, female, white, and enrolled part-time, and more likely to receive financial aid. In addition, the online students tended, in general, to have been better academically prepared, to have come from families of higher socioeconomic status, to have worked more hours for pay while taking courses, and to have been “dual enrolled” before entering college. The outcomes examined in this study are students’ final course grades and their likelihood of withdrawing from a course.

Xu and Jaggars compare the two formats using different statistical models. First, they use a basic ordinary least squares (OLS) model, controlling for a number of student-level variables (such as demographic variables, academic preparation level, and semester-specific information like the number of hours the student was working for pay in a particular semester) and for fixed effects for the semester, the subject of the course, and the campus. In this model, Xu and Jaggars also cluster standard errors both at the student level and at the course level. They find that students in online courses were about 4.4 percentage points more likely to withdraw relative to students in face-to-face courses, and among the students who finished the course, online students tended to earn grades that were about 0.3 point lower than the grades earned by face-to-face students. In addition, Xu and Jaggars use an instrumental variable (IV) approach in which the distance between a student’s home and his college campus serves as an instrument for that student’s likelihood of choosing to take an online course rather than the face-to-face version of that course. The IV approach results in coefficients with greater magnitudes than the OLS estimates, suggesting that the latter may understate the differences between the two formats.

Xu and Jaggars also look at whether the difference between the online format and the face-to-face format vary with certain student characteristics. They find that the males taking online courses were significantly more likely to drop out than were the females taking online courses, but that both genders faced equivalent “penalties” with respect to earning lower grades in online courses. Younger students (defined as students below age 25) also faced significantly larger penalties with respect to grades and withdrawal rates than did older students. By contrast, when prior GPA and credits taken are controlled for, older students appear to do just as well in online classes as in face-to-face classes. Finally, students with lower GPAs (defined as GPAs below 3.0) were significantly more likely to drop out, and earned significantly lower grades, compared to students with higher GPAs.

Relative to much of the other literature, this study includes a large number of controls and involves fairly sophisticated statistical analysis. However, there was no random assignment of participants to one format or the other. In addition, while Xu and Jaggars attempt to use an IV approach to control for unobservable as well as observable differences between students in the two formats, the extent to which they succeed in doing so depends, on part, on the appropriateness of the instrument chosen. The researchers identify three concerns with using distance between a student’s home and campus as an instrument for the likelihood of that student’s enrolling in an online course. However, some of the assumptions they rely on to allay these concerns are arguably questionable—leaving aside the question of whether all the concerns have been identified. For instance, the

researchers assume that proximity to college has no direct effect on the outcome variables, which may not be true if students who live closer to campus have better access to the facilities and instructors they need in order to do well in the course. Xu and Jaggars rule out this possibility based on an analysis limited to enrollments in a subsample of face-to-face courses, in which they find no relationship between distance and outcomes in these courses. This conclusion may not be warranted, however, if the ways in which students in online courses make use of on-campus instructors and facilities differ substantially from the ways students in face-to-face courses make use of these resources.

Finally, and perhaps most importantly, the definition of “online courses” used in this study is very broad, encompassing not only purely online courses but also courses where as much as 49 percent of the instruction took place face-to-face. The researchers also give no indication about what the online courses involved in this study entailed or how much variation existed among those courses (which is unlikely to be negligible, given that about 28,000 of the enrollments they examine were online). Without this information, it is difficult to know the kinds of online courses to which the study’s findings are most applicable, and at most the study shows that the “average” online course in the sample tends to be correlated with higher drop-out rates and lower grades than the “average” face-to-face course in the sample.

Bibliography of Studies

- Ary, Eddie J., and Christopher W. Brune. "A Comparison of Student Learning Outcomes in Traditional and Online Personal Finance Courses." *MERLOT Journal of Online Learning and Teaching* 7.4 (2011): 465-74.
- Bennett, Doris S., Gene L. Padgham, Cynthia S. McCarty, and M. Shawn Carter. "Economics: Internet vs. Traditional Classroom Instruction." *Journal of Economics and Economic Education Research* 8.1 (2007): 21-31.
- Carpinelli, John, Raymond Calluori, Vladimir Briller, Eugene Deess, and Kamal Joshi. "Factors Affecting Student Performance and Satisfaction in Distance Learning Courses." Proceedings of the annual conference of the American Society for Engineering Education, Chicago, Illinois, June 2006.
- Christou, Nicolas, Ivo D. Dinov, and Juana Sanchez. "Design and Evaluation of SOCR Tools for Simulation in Undergraduate Probability and Statistics Courses." Proceedings of the 56th session of the International Statistical Institute meeting, Lisbon, Portugal, August 21-29, 2007. http://www.stat.auckland.ac.nz/~iase/publications/isi56/IPM40_Christou.pdf.
- Colorado Department of Higher Education. "Online Versus Traditional Learning: A Comparison Study of Colorado Community College Science Classes." *WICHE Cooperative for Educational Technologies*. September 2012. Accessed October 22, 2012. wcet.wiche.edu/wcet/docs/blog/1622CCCSOnlinevsTraditionalScienceStudyReportJune2012update.docx.
- Daymount, Thomas, and Gary Blau. "Student Performance in Online and Traditional Sections of an Undergraduate Management Course." *Journal of Behavioral and Applied Management* 9.3 (2008): 275-94.
- Dell, Cindy Ann, Christy Low, and Jeanine F. Wilker. "Comparing Student Achievement in Online and Face-to-Face Class Formats." *MERLOT Journal of Online Learning and Teaching* 6.1 (2010): 30-42.
- Dinov, Ivo, and Juana Sanchez. "Assessment of the Pedagogical Utilization of the Statistics Online Computational Resource in Introductory Probability Courses: A Quasi-Experiment." Proceedings of the Association for Statistics Education, ICOTS-7, July 2-7, 2006, Salvador, Brazil. <http://www.stat.auckland.ac.nz/~iase/publications/17/C312.pdf>.
- Driscoll, Adam, Karl Jicha, Andrea N. Hunt, Lisa Tichavsky, and Gretchen Thompson. "Can Online Courses Deliver In-Class Results? A Comparison of Student Performance and Satisfaction in an Online Versus a Face-to-Face Introductory Sociology Course." *Teaching Sociology* 40.4 (2012): 312-31. doi: 10.1177/0092055X12446624.
- Enriquez, Amelito. "Assessing the Effectiveness of Synchronous Content Delivery in an Online Introductory Circuits Analysis Course." Proceedings of the annual conference of the American Society for Engineering Education, Louisville, Kentucky, June 2010.

Friday, Earnest, Shawnta S. Friday-Stroud, Anna L. Green, and Aretha Y. Hill. "A Multi-Semester Comparison of Student Performance between Multiple Traditional and Online Sections of Two Management Courses." *Journal of Behavioral and Applied Management* 8.1 (2006): 66-81.

Holdhusen, Mark. "Comparison of Engineering Graphics Courses Delivered via Face-to-Face, Online, Synchronous Distance Education, and Hybrid Formats." Proceedings of the annual meeting of the American Society for Engineering Education, Austin, Texas, June 14-17, 2009.

Lewis, J. Scott, and Marissa A. Harrison. "Online Delivery as a Course Adjunct Promotes Active Learning and Student Success." *Teaching of Psychology* 39.1 (2012): 72-76. doi: 10.1777/0098628311430641.

Mentzer, Gale A., John Robert Cryan, and Berhane Teclehaimanot. "Two Peas in a Pod? A Comparison of Face-to-Face and Web-Based Classrooms." *Journal of Technology and Teacher Education* 15.2 (2007): 233-46.

Parsons-Pollard, Nicolle, Robyn Diehl Lacks, and Patricia Hylton Grant. "A Comparative Assessment of Student Learning Outcomes in Large Online and Traditional Campus-Based Introduction to Criminal Justice Courses." *Criminal Justice Studies* 21.3 (2008): 239-51. doi: 10.1080/14786010802355826.

Plumb, Carolyn, and Brock LaMeres. "Comparing Student Learning in a Required Electrical Engineering Undergraduate Course: Traditional Face-to-Face vs. Online." Paper presented at the annual conference of the International Network for Engineering Education and Research, Belfast, Northern Ireland, August 21-26, 2011. http://www.coe.montana.edu/ee/lameres/vitae/publications/d_conference_full/conf_full_024_comparing_live_vs_online.pdf.

Poirier, Christopher R., and Robert S. Feldman. "Teaching in Cyberspace: Online Versus Traditional Instruction Using a Waiting-List Experimental Design." *Teaching of Psychology* 31.1 (2004): 59-62. doi: 10.1207/s15328023top3101_11.

Reasons, Saxon G., Kevin Valadares, and Michael Slavkin. "Questioning the Hybrid Model: Student Outcomes in Different Course Formats." *Journal of Asynchronous Learning Networks* 9.1 (2005): 83-94.

Rich, Anne J., and Mary I. Dereshiwsy. "Assessing the Comparative Effectiveness of Teaching Undergraduate Intermediate Accounting in the Online Classroom Format." *Journal of College Teaching & Learning* 8.9 (2011): 19-28.

Riffell, Samuel, and Duncan Sibley. "Using Web-Based Instruction to Improve Large Undergraduate Biology Courses: An Evaluation of a Hybrid Course Format." *Computers & Education* 44.3 (2005): 217-35. doi: 0.1016/j.compedu.2004.01.005.

Samman, Elfreda, Adnan Omar, Rachid Belmasrou, and David Alijani. "Strategic Plan for Enhancing Online Learning." Proceedings of the Information Systems Educators 2012 Conference, New Orleans, Louisiana, 2012. <http://proc.isecon.org/2012/pdf/1916.pdf>.

Schwartz, Donald A. "Effectiveness of Learning in Online Versus On-Campus Accounting Classes: A Comparative Analysis." *Journal of Research in Innovative Teaching* 5.1 (2012): 63-77.

Signor, Luisa. "Virtual Lectures Versus Face-to-Face Lectures: A Four-Year Study Exploring the Impact on Students' Results." In Geoffrey Crisp, Di Thiele, Ingrid Scholten, Sandra Barker, and Judi Baron (eds.), *Interact, Integrate, Impact: Proceedings of the 20th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education*, Adelaide, Australia, December 7-10, 2003.

Summers, Jessica J., Alexander Waigandt, and Tiffany A. Whittaker. "A Comparison of Student Achievement and Satisfaction in an Online Versus a Traditional Face-to-Face Statistics Class." *Innovative Higher Education* 29.3 (2005): 233-50. doi: 10.1007/s10755-005-1938-x.

Urtel, Mark G. "Assessing Academic Performance Between Traditional and Distance Education Course Formats." *Educational Technology & Society*, 11.1 (2008): 322-30.

Wagner, Suzanne C., Sheryl J. Garippo, and Petter Lovaas. "A Longitudinal Comparison of Online Versus Traditional Instruction." *MERLOT Journal of Online Learning and Teaching* 7.1 (2011): 30-42.

Weber, J. Michael, and Ron Lennon. "Multi-Course Comparison of Traditional Versus Web-Based Course Delivery Systems." *The Journal of Educators Online* 4.2 (2007): 1-19. [www.thejeo.com/Archives/Volume4Number2/Weber Final.pdf](http://www.thejeo.com/Archives/Volume4Number2/Weber%20Final.pdf).

Wilson, Dawn, and David Allen. "Success Rates of Online Versus Traditional College Students." *Research in Higher Education Journal* 14 (2011): 1-8, <http://www.aabri.com/manuscripts/11761.pdf>.

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. Community College Research Center, Teachers College, Columbia University, 2011. www.aefpweb.org/sites/default/files/webform/Online_learning_using_instrumental_variable_approach.pdf.