

COMPARING OUTCOMES BETWEEN ONLINE AND FACE-TO-FACE STATISTICS COURSES: A SYSTEMATIC REVIEW

Kathleen Mathieson

Department of Interdisciplinary Health Sciences, A.T. Still University, United States of America
kmathieson@atsu.edu

The increase in online learning has led to research comparing student outcomes in online and face-to-face (FTF) statistics courses. This systematic review presents an analysis of studies comparing student achievement and student satisfaction in statistics courses taught online to those taught FTF. Multiple databases, conference programs, and bibliographies from selected articles were searched. Among the 411 studies identified, 13 met inclusion criteria. Student achievement was generally similar between online and FTF methods, but results on student satisfaction were inconclusive. Studies comparing online and FTF statistics courses are characterized by considerable methodological limitations. Instructors teaching statistics online would benefit from future research that is grounded in sound methodology, that is informed by past research and theory, and that examines specific ways to improve learning in an online environment.

INTRODUCTION

The dramatic increase in online instruction over the past few decades (Allen & Seaman, 2007) has led to discussion, research, and debate regarding the efficacy of online instruction relative to traditional, face-to-face (FTF) instruction. Some argue that online and FTF instruction have comparable outcomes (e.g., Russell, 1999). Critical analyses of the evidence, however, result in a less definitive conclusion because studies often lack the methodological rigor necessary for establishing a true cause-and-effect relationship (Cook et al., 2008; Phipps & Merisotis, 1999). This paper presents a systematic review of such research and discusses the results in the context of the broader literature on statistics education.

METHOD

Search Strategy

For inclusion, three criteria had to be met. First, studies must have compared statistics courses taught primarily online (i.e., via the internet) to FTF statistics courses. Studies comparing hybrid courses (part online, part FTF) and those comparing FTF with other distance methods (such as courses delivered via CD-ROM or two-way television) were excluded. Second, studies must have reported outcome data on student achievement (e.g., course grades, exam scores) and/or satisfaction (e.g., satisfaction survey, course evaluations). Third, studies must have been conducted in a postsecondary setting. Since the World Wide Web was introduced in 1991, the search was limited to studies published or presented after 1990.

To identify potentially relevant studies, first the following databases were searched: ERIC, PsycINFO, CINAHL and ProQuest Digital Dissertations. Second, the following journals were searched: Journal of Statistics Education, Statistics Education Research Journal, and Journal of Online Teaching and Learning. Third, conference proceedings for the following organizations were searched: American Statistical Association, International Association for Statistics Education, International Statistical Institute, and American Psychological Association. Finally, the literature index of the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) Web site was searched. Full text copies of studies that appeared to meet inclusion criteria were retrieved and reviewed to determine final inclusion status. Finally, references of articles meeting inclusion criteria were hand-searched to identify any additional studies not discovered in the initial search process.

Review Process

Study quality was determined by rating each study according to its level of evidence. Levels of evidence are commonly used in the health professions to assess the quality and rigor of research studies, to help guide interpretation, and to inform evidence-based practice. The

guidelines used to assess level of evidence in this study are described in Lieberman and Scheer (2002). This framework assesses study quality based on design, sample size, and internal and external validity. Data such as student population, discipline, and outcomes were abstracted.

RESULTS

Out of 411 studies identified during the search process, about 120 were retrieved for full text review, and 13 of these met the inclusion criteria. Many case studies of instructors' personal experiences with teaching online statistics courses were excluded because they did not include a comparison group.

Most of the 13 studies were conducted in the United States, with the exception of two conducted in Israel (Katz & Yablon, 2003; Yablon & Katz, 2001) and one conducted in Thailand (Suanpang & Petocz, 2006). Studies included graduate and undergraduate students from a range of disciplines, including business, nursing, public health, and social sciences. Levels of evidence and findings of the selected studies are detailed in Table 1.

Table 1. Summary of studies included in review

Study	Level of Evidence*	Student Outcomes	
		Achievement	Satisfaction
Dutton & Dutton (2005)	IIA2b	Online Higher	Not Assessed
Evans et al. (2007)	IIB2b	No Difference	Not Assessed
Gunnarsson (2001)	IIB2b	No Difference	Not Assessed
Hurlburt (2001)	IIA3b	No Difference	Not Assessed
Jones (1999)	IIA2b	No Difference	Not Assessed
Katz & Yablon (2003)	IIA3b	Online Higher	Online Higher
McGready (2006)	IIA2b	No Difference	Not Assessed
McLaren (2004)	IIA3b	No Difference	Not Assessed
Schutte (1998)	IB3b	Online Higher	Online Higher
Suanpang & Petocz (2006)	IA2c	Online Higher	Not Assessed
Summers et al. (2005)	IIB3b	No Difference	FTF Higher
Wang & Newlin (2000)	IIA2b	FTF Higher	Not Assessed
Yablon & Katz (2001)	IIA3b	No Difference	Online Higher

*Level of evidence is determined by quality of study design (I, II, III, or IV), sample size (A if ≥ 20 per condition, B if ≤ 20 per condition), internal validity (1, 2, or 3), and external validity (a, b, or c). For example, a randomized, controlled trial with a large sample size and high internal and external validity would receive a rating of IA1a.

Study Quality and Level of Evidence

Only two of the 13 studies (Schutte, 1998; Suanpang & Petocz, 2006) used an experimental design in which students were randomized into online or FTF courses, and among the 11 lacking random assignment, four did not test for baseline differences between study groups (Hurlburt, 2001; Katz & Yablon, 2003; McLaren, 2004; Yablon & Katz, 2001). Given these limitations, and since students registering for online courses often have different sociodemographic and academic characteristics from those registering for FTF courses (Cercone, 2008), none of the studies in this review met the highest level of evidence for internal validity; there were often alternate explanations for the outcomes due to failure to fully account for group differences on baseline characteristics. Ten studies included at least 20 students per group, thus meeting the highest level of evidence for sample size. None of the studies were rated as having high external validity due to unique samples that limited generalizability of results.

Student Achievement

Eight studies assessing student achievement found no significant difference between online and FTF students. In the five studies finding a difference, four reported higher achievement in the online group and one reported higher achievement in the FTF group (see Table 1).

Student Satisfaction

Only four studies compared student attitudes. Among these, three reported higher satisfaction in the online course, and one reported higher satisfaction in the FTF course. However, all of these studies received the lowest score for internal validity due to a failure to randomize students to groups, a failure to adjust for baseline differences, or fundamental differences in instruction between the online and FTF courses.

CONCLUSION

Most of the studies comparing statistics courses taught online to those taught FTF found student achievement to be comparable between modes. Although this finding is consistent with some recent meta-analyses (Cook et al., 2008; Schenker, 2007), it is difficult to generalize findings based on course exams or grades since the course assessments used are often unique to each study, poorly defined, and do not have established reliability or validity (Zieffler et al., 2008). Firm conclusions cannot be drawn regarding student satisfaction due to the small number of studies addressing satisfaction and the lack of internal validity in those studies.

The limitations of educational research in general (Hargreaves, 2007) also apply to research examining distance versus FTF instructional methods. As Phipps and Merisotis (1999) state, "The most significant problem is that the overall quality of the original research is questionable and thereby renders many of the findings inconclusive" (p. 3). Perhaps the most notable finding of this review is the prevalence of methodological limitations in the reviewed studies, such as lack of randomization, lack of generalizability, differences in baseline characteristics between groups, and failure to adjust for confounders. Indeed, a recent review (Zieffler et al., 2008) discusses limitations of studies comparing instructional methods in college statistics courses and recommends that researchers use valid and reliable measures for concepts such as achievement, and take steps to enhance internal and external validity.

Some argue that comparing online and FTF courses is no longer useful and that future research should examine specific methods of online instruction (Cook et al., 2008) or attributes of successful online learners (Artino, 2008). Online instruction certainly shows no sign of decreasing and instructors are often charged with the task of developing and/or teaching online statistics courses to serve certain students (Summers et al., 2005). In those cases, the relevant question is not whether online instruction is comparable to FTF instruction, but how to make it effective in its own right. Situating such work in the context of the literature on statistics education is essential. A good example is Everson and Garfield's (2008) review of how the Guidelines for Assessment and Instruction in Statistics Education (GAISE) can be implemented into the discussion component of online statistics courses. Similar research examining specific ways to improve learning in online statistics courses, grounded in sound methodology and informed by past research and theory, has the potential to be very useful to instructors who are teaching statistics in an online environment.

REFERENCES

(*Included in systematic review)

- Allen, I. E., & Seaman, J. (2007). *Online nation: Five years of growth in online learning*. Retrieved October 23, 2009 from http://sloanconsortium.org/publications/survey/pdf/online_nation.pdf.
- Artino, A. R. (2008). Promoting academic motivation and self-regulation: Practical guidelines for online instructors. *TechTrends: Linking Research and Practice to Improve Learning*, 52(3), 37-45.
- Cercone, K. (2008). Characteristics of adult learners with implications for online learning design. *AACE Journal*, 16(2), 137-159.
- Cook, D. A., Levinson, A. J., Garside, S., Dupras, D. M., Erwin, P. F., & Montori, V. M. (2008). Internet-based learning in the health professions: A meta-analysis. *Journal of the American Medical Association*, 300(10).
- *Dutton, J., & Dutton, M. (2005). Characteristics and performance of students in an online section of business statistics. *Journal of Statistics Education*, 13(3). Retrieved October 23, 2009 from <http://www.amstat.org/publications/jse/v13n3/dutton.html>

- *Evans, S. R., Wang, R., Yeh, T., Anderson, J., Haija, R., McBratney-Owen, P., Peeples, L., Sinha, S., Xanthakis, V., Rajcic, N., & Zhang, J. (2007). Evaluation of distance learning in an "Introduction to Biostatistics" class: A case study. *Statistics Education Research Journal*, 6(2), 59-77.
- Everson, M. G., & Garfield, J. (2008). An innovative approach to teaching online statistics courses. *Technology innovations in statistics education*, 2(1). Online: <http://repositories.cdlib.org/uclastat/cts/tise/vol2/iss1/art3>
- *Gunnarsson, C. L. (2001). *Student attitude and achievement in an online graduate statistics course* (Doctoral dissertation). Retrieved from ProQuest Digital Dissertations. (AAT 3000364).
- Hargreaves, D. H. (2007). Teaching as a research-based profession: Possibilities and prospects. In M. Hammersley (Ed.), *Educational research and evidence-based practice* (pp. 3-17). London: Sage.
- *Hurlburt, R. T. (2001). "Lectlets" deliver content at a distance: Introductory Statistics as a case study. *Teaching of Psychology*, 28(1), 15-20.
- Phipps, R. A., & Merisotis, J. P. (1999). *What's the difference? A review of contemporary research on the effectiveness of distance learning in higher education*. Washington, DC: American Federation of Teachers and National Education Association.
- *Jones, E. R. (1999). A comparison of an all web-based class to a traditional class. Paper presented at the *Society for Information Technology & Teacher Education International Conference, San Antonio, TX*.
- *Katz, Y. J., & Yablon, Y. B. (2003). *Locus of control, self-esteem, motivation and satisfaction, teaching and learning through an internet-based and traditional "Introduction to Statistics" course*. *Proceedings of European Distance Education Network* (pp. 183-197). Rhodes, Greece: EDEN.
- Lieberman, D., & Scheer, J. (2002). AOTA's evidence-based literature review project: An overview. *American Journal of Occupational Therapy*, 56(3), 344-349.
- *McGready, J. (2006). Basic biostats: Online learning versus onsite learning. Paper presented at the *Joint Statistical Meetings, Seattle, WA*.
- *McLaren, C. H. (2004). A comparison of student persistence and performance in online and classroom business statistics experiences. *Decision Sciences Journal of Innovative Education*, 2(1), 1-10.
- Russell, T. (1999). *The no significant difference phenomenon*. Raleigh, NC: North Carolina State University.
- Schenker, J. D. (2007). The effectiveness of technology use in statistics instruction in higher education: A meta-analysis using hierarchical linear modeling. (Doctoral dissertation). Retrieved from ProQuest Digital Dissertations. (AAT 3286857).
- *Schutte, J. G. (1996). *Virtual teaching in higher education: The new intellectual superhighway or just another traffic jam?* Unpublished manuscript, Department of Sociology, California State University, Northridge, CA. Retrieved October 23, 2009 from <http://www.csun.edu/sociology/virexp.htm>.
- *Suanpang, P., & Petocz, P. (2006). E-learning in Thailand: An analysis and case study. *International Journal of E-Learning*, 5(3), 415-438.
- *Summers, J. J., Waigandt, A., & Whittaker, T. A. (2005). A comparison of student achievement and satisfaction in an online versus a traditional face-to-face statistics class. *Innovative Higher Education*, 29(3), 233-250.
- *Wang, A. Y., & Newlin, M. H. (2000). Characteristics of students who enroll and succeed in psychology web-based classes. *Journal of Educational Psychology*, 92(1), 137-143.
- *Yablon, Y. B., & Katz, Y. J. (2001). Statistics through the medium of the internet: What students think and achieve. *Academic Exchange Quarterly*, 5(4), 17-22.
- Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., & Chang, B. (2008). What does research suggest about the teaching and learning of introductory statistics at the college level? A review of the literature. *Journal of Statistics Education*, 16(2). Online: www.amstat.org/publications/jse/v16n2/zieffler.html.