Traditional, Web-based, and Hybrid Instruction: A Comparison of Training Methods

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Richard Nathaniel Landers

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Adviser: Paul R. Sackett

May 2009

Acknowledgements

I was always told that a surprisingly large number of people go into the making of any particular Ph.D., and upon reflection, I find that I am no exception. My original interest in computers and technology comes from my father, Ric Landers, who when I was only a few years old, made my name flash in a variety of colors in a BASIC program on our IBM PCjr. Since then, thanks to him, I've been hooked on technology. His constant support in whatever turn I decided my education would take since then has been more appreciated than he will probably ever know. I'm also very thankful for my grandmother and late great grandmother, who always showed a great deal of support even if they weren't ever quite sure what I was doing or why I was doing it so many miles from home.

Until college, I was all set to be a computer programmer until a graduate student instructor at the University of Tennessee at Knoxville whose name I don't remember gave me a very interesting introduction to psychology. After that, I sampled upper division classes, trying to figure out what area sounded the most interesting, until taking Dr. John Lounsbury's class on Testing in Psychology, where he enthusiastically introduced me to the ideas of trait personality and quantitative measurement. As John's undergraduate advisee, I took a course in organizational psychology with Dr. Eric Sundstrom, cementing my future in I/O. Dr. Lounsbury also encouraged me to push myself as an undergraduate, ultimately leading to my first publication in *Computers in Human Behavior*. This, along with his encouragement to strive for the best I could achieve, ultimately led me to the University of Minnesota for graduate school. I wouldn't be where I am today without his influence, and I am very grateful for all the guidance he provided.

At the U of M, I have been truly fortunate to work with some of the greats in I/O psychology. Drs. John Campbell, Deniz Ones, and Nathan Kuncel, aside from taking time from their busy schedules to serve on my dissertation committee, provided valuable instruction throughout my time as a graduate student, both in terms of subject matter expertise and as models of what I/O psychologists should be. I am also very grateful to Dr. Connie Wanberg for taking several long trips across the river to help shape my dissertation, Dr. Joyce Bono for the instruction and work opportunities that she provided, and Dr. Jim Butcher for the professional opportunities that he provided.

Of all the faculty at Minnesota, I am most and eternally grateful to my adviser, Paul Sackett. His willingness to entertain any research idea, discuss any problem, and return any dissertation draft with astonishing speed has given me a rock-solid foundation on which to build my career, and I could not have been more fortunate. I credit him as the base on which all of my successes as a graduate student stand and cannot imagine how I would have made it through graduate school with anyone else in my corner. He is the model on which I plan to base my own advising, and it's a very high bar to reach for.

Of course, faculty alone did not provide all of the support that I received at the U; I am quite thankful toward all the I/O current and former graduate students that I've

interacted with. My increasingly shrinking cohort in particular – ultimately consisting of Brian Connelly and Matt Borneman – provided a constant source of distraction, friendship, and camaraderie for which I am very grateful. I am also especially thankful to Stacy Eitel Davies for her generosity during the final stages of my dissertation and job search, as her willingness to teach an astonishing seven classes in my stead was above and beyond the call of duty.

Outside of the U, I am grateful to Don MacPherson, Dan Riley and Pat Riley at Modern Survey for providing me with summer internships during every year of graduate school, as well as a place to see how my academic work in I/O relates to "the real world." Don in particular provided friendship and personal guidance that will stay with me for a long time. I am also thankful to Drs. Don Davis, Debbie Major, and Karin Orvis at Old Dominion University for their enthusiastic support in welcoming me into their department.

Last, and certainly not least, I am thankful to my wife, Amy. We met in our graduate school statistics class, got to know each other through the shared trials of seeking the doctorate, were married in Kyoto between our fourth and fifth years, and are now moving away from the icy north to start our "real" life together. Throughout graduate school and my dissertation in particular, she provided encouragement when I needed to be pushed, support when I needed stability, distractions when I needed to relax, and humor when I needed to smile. I cannot imagine a life without her, and I feel unbelievably lucky that we found each other when we did, even if we were buried in

ANOVAs at the time. I doubt I will ever be able to show her how happy I really am for her to be a part of my life, and a few sentences between pages iii and iv hardly seem like enough.

So, as the saying goes, it takes a village, and I have been very lucky to have a very supportive village. The problem with acknowledgements is that the village roster is always longer than the pages, and I am sure that I am missing at least a few people absolutely vital to my success. So to all those listed and those that should have been, thank you all for being a part of my life and my career. I hope to be lucky enough to know you all for a long, long time.

Abstract

Previous efforts to summarize differences between traditional and web-based training program learning outcomes have been limited by ignoring pre-training differences due to non-random assignment to conditions. In the present meta-analysis, pre-training and post-training differences between media (traditional, web-facilitated, hybrid, and fully online instruction) were summarized across a variety of moderators. Self-selection into condition (e.g. college students choosing to take a traditional or online course according to their preferences) lead to substantial outcome differences before training begins; trainees choosing online courses know more (d = .19) about the material on which they are going to be trained than those choosing traditional courses. This makes interpretation of post-training outcome differences alone unwise.

To address this, three approaches were taken: 1) pre-training and post-training between-subject *d*'s (traditional vs. web-based instructions) were compared, 2) traditional and web-based within-subject *d*'s (post-training vs. pre-training scores) were compared, and 3) between-subject post-training *d*'s adjusted by pre-training differences for those studies in which self-selection was present were examined.

These analyses led to several conclusions, including 1) within-subject outcome gains are large in comparison to between-subject differences in media, 2) differences do exist across criterion type (knowledge, observable skill, problem solving skill, attitudes, and perception of gain), 3) the effectiveness of web-based courses in comparison to traditional courses appears to be changing over time, perhaps as the design of online courses becomes more sophisticated, 4) the use of learner control in online courses appears to produce worse outcomes for learners, 5) online courses produce better outcomes than traditional courses when those courses take place over a longer time period, and 6) learners self-selecting into online courses seem to be older and have greater computer/Internet experience than those self-selecting into traditional courses. Areas of e-learning research needing more primary empirical research are identified, and suggestions for training design practitioners are discussed.

Table of Contents

List of Tables

Introduction

Electronic learning (e-learning) is the likely future of most formal work-related training. It already holds a strong presence in higher education and continues to grow, with an estimated 3.1 million students taking at least one online course during Fall 2005, an increase of 0.9 million students from the previous year. From 2002 to 2004, enrollment in online courses increased by approximately 20% each year, with no plateau yet in sight. This instruction occurred at all levels of higher education, including associate, baccalaureate, master's, and doctoral-level programs (Allen & Seaman, 2006). Specifics numbers of learners are not available for corporate settings, but a report from market research organization Global Industry Analysis revealed that the corporate e-learning market in 2007 was $17.5 billion, with a projected growth to $52.6 billion by 2010 (Kopf, 2006). There is clearly enthusiasm for training supported by computers, and the ever-evolving Internet draws a great deal of interest in particular. It is unfortunate then, that we actually know relatively little about the effectiveness of online training in comparison to more traditional methods. Studies examining such differences in media (for example, comparing an in-person face-to-face training course directly with a fully online course) are most often poorly designed with unconsidered study-level characteristics influencing their results, making it difficult to determine what the true differences between media are in terms of both learning and reactions to training. It is the central purpose of this thesis to investigate these very concepts. Thus, the overall question addressed here is: "Are there overall differences in effectiveness and reactions between face-to-face instruction and web-based learning, and what other factors might moderate this difference?"

Part of the reluctance to investigate these systems in psychology may stem from the fear that the movement of training into the electronic space may be nothing more than a fad in the classic sense (Dunnette, 1966). E-learning is by its nature filled with buzzwords that warn of faddishness. Purveyors of new technologies typically name them so that they are memorable, and it is often difficult for the casual observer to distinguish between serious technologies and passing fads. Are Twitter and Facebook resources to consider for online training? Are blogs, wikis and podcasts tools to stay? Or will money invested in designing training relying on these technologies prove wasted as they fade away into technological history? Many terms are even invented to address the same concept with different levels of specificity; for example, computer-based training, learning management systems, distance learning, virtual learning environments, and e-learning all describe the same general category of tools, with relatively minor differences. To ensure clarity of the discussions herein, I will first set up a basic terminology by which to consider these technologies while simultaneously describing the history of these terms.

Terminology and a Brief History of Web-based Training

From the broadest perspective, the topic of this thesis is employee training, a systematic process by which employees gain new knowledge, skills and attitudes based upon organizational needs (Goldstein & Ford, 2002). In the modern workplace, such formal training most often takes place in the form of in-person instructor-led classes (61.2%; American Society for Training and Development, 2008): a training designer or other expert stands at the front of a group of trainees and delivers training content, most often verbally, although many times with the assistance of a PowerPoint presentation or paper handouts. Some interaction might be present; such training sessions are often called “workshops.” I will hereafter refer to this model as “traditional” training. Unfortunately, traditional training is somewhat limited in that a training expert must be physically present in the same room as the trainees in order to deliver content. There are two categories of situations in which this is a hindrance to an organization. First, if expertise is difficult to find, or alternately, if a limited number of experts are available, then the time of those experts becomes increasingly valuable and difficult to schedule. For example, if a large multinational developer of laptop computers releases a new model with many new features to potentially thousands of retail locations, it is not feasible (or at least, cost-effective) to send experts on the new product from corporate headquarters to every retail location in which it is to be sold. Second, if the trainees themselves are not colocated, the problem is exacerbated. For example, consider a large real-estate firm in which most agents employed by the company work by remote. These individuals schedule meetings with their own clients in a variety of locations across a wide physical area and may rarely come to the central office. Requiring all of those agents to meet at a central location at a single point in time for a one hour lecture on new sales regulations may damage not only their individual sales (decreasing job performance) but those agents' willingness to work with a firm taking their time unnecessarily (increasing turnover).

To combat these difficulties, distance learning evolved out of the traditional training model. Instead of bringing trainees to one location to receive training material, the training material was brought to them. This has taken many forms since its inception, but generally falls fairly close to the traditional model of instruction. For example, lecture-by-videotape, in which a training designer records their lectures and mails them to trainees to watch on their own time, is a common form of distance learning. Although this form of distance learning may be transitioning to DVD or other such media, the general principle is the same: a lecture as it would be given traditionally is recorded and played back for trainees. As computers grew in popularity during the late 1980s and 1990s so did computer-based training, the definition of which is as straightforward as it sounds: training received on a computer. Text-filled documents would be placed on a floppy disk which would be delivered to trainees at a distance for them to read at their convenience. As the technology evolved, this soon grew to include graphics (i.e. pictures and drawings), and eventually audio and video. Simultaneously, the media used evolved from disks to CD-ROM to transfer via the Internet.

The introduction of the Internet to the training world in the early 2000s also introduced new possibilities. Before this time, a course or training program existing in a lecture hall and converted to a computer-based training program changed little in terms of content. While the specific words heard or read using either method might differ, the overall content and instructional method usually did not. For example, if the in-person instruction was simply a lecturer standing at the front of a group of trainees verbally delivering content, then the computer-based material was likely static text on the screen or a video of a lecturer delivering that same content. Many of these design choices were made because of the relative inflexibility of the technology. In the 1990s, despite the fact that a computer-based training module with a great deal of interaction and engagement with the learner was possible, such interactive modules were very expensive and unlikely to be developed quickly enough for the training it presented to be relevant by the time it was complete.

With the explosion in popularity of the Internet in the early 2000s, it became increasingly important that when trainers delivered content via the Internet, and the World Wide Web in particular, they could be reasonably confident that the user experience would be similar regardless of the computer that the user was using. This standardization effort from the World Wide Web Consortium (W3C) drove development costs down for all developers of projects on the Web, making innovative training projects on the Web far simpler and less expensive to attempt.

With decreased costs, increased production on the web across all industries soon followed, making the Internet a forum packed with information for the modern consumer. Demands for access to the Internet increased rapidly in response, and by 2003, over half of U.S. households (over 61 million) had Internet access in their home, up 50% from just two years earlier (U.S. Census Bureau, 2005), with continued growth projected. This growth was not limited to home use; those that access the Internet at work spent a mean of 17.8 hours per week accessing the Web while at work in December 2008 (Nielsen Company, 2009).

With decreased development costs and Internet access available to the majority of Americans, the movement of training and instructional programs to the Web picked up speed shortly thereafter. Fully web-based (or “online”) courses were created with multiple objectives in mind: decreasing costs and increasing accessibility. These courses most often still followed the traditional model of instruction; instead of accessing non-interactive text and videos via a CD-ROM, students and trainees would access this material via a web browser.

Another variety of course or training program also appeared as a blend of the traditional and web-based models of instruction. These hybrid or blended courses come in many forms, and are inconsistently labeled across studies. The Sloan Consortium, the foremost non-profit organization dedicated to improving the quality of distance education, splits courses into four categories. First, those with 0% of content delivered online are referred to as “traditional” courses. Second, those with 1 to 29% of content delivered online are referred to as “web-facilitated” courses. These courses are primarily given face-to-face but incorporate some aspects of the Web to support course activities. Course management systems (sometimes called learning management systems) are the hallmark of this type of course; these are online hubs where learners and instructors coordinate. They are typically able to give access to course material (such as a syllabus, video), group discussion tools (such as real-time chat or discussion boards), and online assessments, although the instructor has the freedom to include whichever specific course components she wishes. Modern examples of course management systems are Blackboard, WebVista, and Moodle. Third in the list of course types are “blended/hybrid” courses, in which 30 to 79% of material is delivered online. Finally, fully “online” courses are those in which 80% or more material is delivered online (Allen & Seaman, 2006). This categorical split is somewhat arbitrary. There are actually no specific points at which a course switches from one type to another, and these labels are used primarily for convenience.

Effectiveness of Distance Learning Courses

A major concern early in distance learning research was that of effectiveness. Could a course given at a distance be as effective as one taught in-person? One of the oldest distance learning techniques, the audio-tutorial (A-T) approach, was the focus of much scrutiny in this regard. Created by biologist Samuel Postlethwait in 1961, the A-T approach involved the delivery of a course via audiotape recordings, either in conjunction with an in-person class or as a standalone mini-class (Postlethwait, Novak, & Murry, 1971). With the invention of the compact cassette tape in 1963, this became an increasingly attractive option for those instructors wishing their distance learning courses to be more engaging and effective than one delivered on paper alone. Concerned with the question of instructional effectiveness, Postlethwait in 1962 conducted one of the earliest “media comparison studies” by assigning an experimental section of a course to receive instruction entirely by the A-T approach, and the control group receiving it in a more traditional lecture. He reported no differences in outcomes between the two conditions.

As the A-T approach grew in popularity, two efforts to summarize the growing research literature appeared at roughly the same time (J. A. Kulik, Kulik, & Cohen, 1979). In 1975, the first review to appear was by Mintzes, who summarized six studies regarding A-T in collegiate settings and concluded that the literature examining the effectiveness of the A-T approach was inconclusive. Three studies that Mintzes examined showed A-T to be superior to traditional instruction, two showed no differences, and one showed traditional instruction to be superior. The second and a more expansive review by Fisher and MacWhinney (1976) appeared in the following year, concluding strongly that the A-T approach was indeed more effective in collegiate, high school, and elementary school settings. Eighteen of forty-four studies showed A-T to be superior, twenty-five showed it to be equivalent, and only one showed it to be inferior.

Each of these reviews relied on counting as its method of aggregating study results. By counting the number of statistically significant findings compared to the total number of studies available, each of these authors hoped to gain a sense of the overall effectiveness of A-T. However, conflicting conclusions between the two reviews created a number of questions. Is A-T stronger in high school and elementary settings than in colleges? Is sampling error making the results of a six-study review more variable? Or do the authors simply disagree on their interpretation of essentially the same data? Faced with this ambiguity, J. A. Kulik, Kulik and Cohen (1979) conducted a Glassian meta-analysis, synthesizing the results of 48 studies on college students comparing the A-T approach with traditional models of instruction. They examined both student achievement (d = .20, in favor of A-T) and student evaluations (d = .12, also in favor of A-T) as outcome variables. With this, they concluded that A-T was indeed superior to traditional instruction, although as suggested by Cohen (1977), the difference was relatively small. C.-L. Kulik, Kulik and Cohen (1980) later expanded this analysis by including more computer-based instructional technologies, increasing the number of analyzed studies to 278, but the findings were similar. A small to moderate overall difference across technologies was observed (d = .28). They also noted that studies using different instructors in the two instructional conditions (technology-enhanced/experimental and traditional/control) exhibited a larger difference (d = .41) than those using the same instructor (d = .20).

The Clark-Kozma Debate

In one of the most influential papers regarding instructional technology, Clark (1983) argued that the effects observed in the Kulik papers were not "instructionally significant and could easily be due to confounding" (p. 448) and similarly condemned all media comparison research. His argument hinged primarily around the instructorship moderator found by C.-L. Kulik, Kulik and Cohen (1980). The differences observed, he argued, were not due to differences in media but instead to a variety of other factors.

First, method and content often varied with medium; in other words, when the instructor varied, the instructors were using different instructional methods (and sometimes different material altogether), and thus medium was not the only difference between conditions. When the instructors were not different, he posited that the remaining smaller difference (*d* = .20) was “due to systematic but uncontrolled differences in content and/or method, contributed unintentionally by different teachers or designers” (Clark, 1983, pp. 448-449). He offered no empirical evidence to support this position, and this position is weakened further with the appearance of work by Kulik, Bangert and Williams (1983). In their meta-analysis of a similar literature, the mean effect size for “different instructor” was .30, while the mean effect mean size for “same instructor” was .28. Second, he continued, the difference could be due to increased effort on the part of the instructor. Because the technology being adopted was new, it required additional “design effort,” presumably creating a superior course. No empirical evidence was provided here either. Third, he argued that the instructional methods being studied were really just that – methods – and were not and should not be tied to the medium in which they were used. Researchers studying instructional media at the time tended to develop their techniques to be used on a computer, even if the technique could be used in person. Thus, the differences observed were not due to medium but again due to instructional method. Fourth, he argued that a novelty effect was artificially increasing observed effect sizes. As support, he cited the findings of Kulik, Bangert and Williams (1983) that studies lasting four weeks or less showed a sizable difference favoring technology (*d* = .56), while the differences in final outcomes of those lasting longer were increasingly smaller (*d* = .30 for 5-8 weeks, and *d* = .20 for more than 8 weeks). He failed to note the rival hypothesis that instructional technologies might in fact be superior in short term training interventions while weaker in longer interventions. Instead, he called this “the familiar .2 effect” (Clark, 1983, p. 450) and dismissed any importance it might hold. His fifth and final criticism was that of a file drawer effect, the tendency for papers reporting stronger results to be published while papers with weaker results to remain unpublished.

With this position laid out, Clark (1983) called for a moratorium on all media comparison research, concluding “five decades of research suggest that there are no learning benefits to be gained from employing different media in instruction” (p. 450) and stating rather famously K (p. 445). This claim slowed all research examining potential media differences until 8 years later when Kozma (1991) directly challenged Clark’s position. Kozma responded by stating that different technologies might possess cognitively relevant characteristics that alter the learner’s formation of mental models. For example, a symbol system such as plain text can be delivered in a variety of formats. A book offers a permanent, fixed source of information to be referenced. A scrolling marquee delivers the same symbol system, but in an unstable format. Permanence of information might be considered a construct underlying observed media differences, and this construct might be the causal mechanism behind differences in media effectiveness. Thus, media differences between a book and scrolling marquee might directly influence a learner's ability to form mental models based upon the information presented, even if the information itself is identical between the two formats.

Thus began the "Clark-Kozma debate." A flurry of writings across several publications appeared, culminating in a 1994 special issue of *Educational Technology Research & Development*. Clark (1994) directly responded to Kozma's assertions by suggesting that "replaceability" was the key to unraveling this issue. He stated "I challenge Robert Kozma and other colleagues in this area to find evidence, in a well designed study, of any instance of a medium or media attributes that are *not* replaceable by a different set of media and attributes to achieve similar learning results for any given student and learning task" (p. 22), suggesting again that method, not medium, is the true causal force eliciting differences in media comparison research. He further equated all empirically observed differences in media with differences in method, stating "when a study demonstrates that media attributes are sufficient to cause learning, the study has failed to control for instructional method and is therefore confounded" (p. 25), placing himself in a position that, logically speaking, was not falsifiable. Kozma (1994b) generally agreed with Clark's overall position, but stated that Clark was only accurate currently because of historical constraints. Kozma argued that the equivalence of media and method in many studies was only due to a lack of innovation. Media *could* be used in a way that would uniquely contribute to learning beyond differences in method alone; it simply was not doing so at this time, or at least not in large quantities, due to a lack of technological sophistication. He criticized Clark's delivery truck metaphor as short-sighted behaviorism, stating “the medium is an inert conveyor of an active stimulus to which the learner makes a behavioral response. However, as we have come to understand…learning is an active, constructive, cognitive and social process” (pp. 8-9). He argued that the interaction between the learner and the learning environment produces knowledge, thus making media key to the process. Clark’s position thus focused on the attributes of media *necessary* to bring about learning. He believed these attributes were always present as long as the method was physically able to convey the material. By this view, all media are essentially equivalent as long as the instructional method behind them is sound, and thus the lowest cost option should always be used. Kozma’s position focused on the attributes of media *sufficient* to bring about learning in as effective a manner as possible. He believed each media brought with it certain advantages and disadvantages when providing instruction, and that these qualities varied depending upon the material to be taught. By this view, media should be chosen when designing or implementing an instructional method so as to maximize the value provided by those media. In fact, as he argued in the following issue, research comparing the best designed instructional programs with previous systems will confound media and method to the maximum possible degree (Kozma, 1994a).

Commentaries on these two articles in the special issue were varied. Jonassen, Campbell and Davison (1994) criticized the debate for being too focused on designing instructional systems from an instructionist perspective. Computer systems, they argued, should be designed to support learners actively and not deliver instruction to them passively. Morrison (1994) argued that Clark's and Kozma's positions were not in fact contrary to one another, or even related closely enough to refer to the same question. Reiser (1994) added that from a practitioner's perspective, Clark's position only holds when using Clark's rather narrow definitions, and that the entire debate is unnecessary to anyone actually working in the field. Shrock (1994) echoed Reiser's concerns about definitional ambiguity and scope as well as Jonassen et al.'s concerns about instructionism. The only aspect of the Clark-Kozma debate agreed upon by all contributors (save perhaps Clark) was that further argument was likely a waste of time. Thus sits the debate today.

Media Comparisons Portrayed in the I/O Psychology Literature

The vast majority of contributions to the media comparison debate, both theoretically and empirically, came from the field of education technology. In business settings, only a few scattered studies have tackled the question, and many exhibit the research design problems originally described by Clark (1983) that plague the educational technology literature. Schmeekle (2000), for example, compared outcomes for jail management trainees randomly assigned to either classroom or online training. No statistically significant difference in knowledge was found between the conditions; however, a difference was found for the amount of time spent in training. The author concluded that the two media were equally effective, but that online training was demonstrably more "efficient" (p. 240) than classroom instruction. Unfortunately, many of the research design concerns that Clark criticized of educational technology research likely exist here as well. Very little information is given on the nature of the training program itself, except that it was designed on a new government grant some time after the in-person instruction had been in use. It was designed explicitly to "provide a more accessible training method for law enforcement personnel, decrease training time, and reducing training costs" (p. 218), and this newly designed system was contrasted with a system in place for some time. Lacking any other information about the training programs, this strongly suggests that much more was varying between conditions other than simply "media," probably including accessibility and the length of the material. No effort was made to describe similarities or dissimilarities between the two programs, a common omission in such research. The only conclusion that should be made from this study is that whatever online training intervention they designed took less time for trainees to complete and was nearly as effective as whatever training intervention they had used previously. It provides no information on the specific mechanisms or components of these differences. Probably the most troubling aspect of this study is that a training designer could easily conclude from it that simply converting a classroom-based course to a web-based course would decrease time-to-completion without any other negative effects on training outcomes. Instead, the specific designs of the two training programs may have played a much larger role in the observed lack of differences.

Because such class-level confounds are common in both the educational and organizational literatures, it is difficult to generalize the results of any single study of a single course or training program to any other training program. Arbaugh (2005) argues that studies based upon single courses, comparisons of online and in-person versions of the same course, and multi-university replications of courses are all of limited use due to such confounds, and extended this advice to comparisons of training programs as well. In order to be tested empirically, most modern theories of successful web-based instruction require multiple training sessions or courses to be run. Learner-participants are not the cases of interest in analysis; instead, courses are. Individual learners are nested within courses, and differences in course outcomes based upon course design provide the data needed to answer interesting questions in this area. Arbaugh (2005) followed his own advice by examining survey and outcome data from 44 sections of the same MBA program. Although not a comparison of in-person and online courses, this is an excellent study design and of much greater value than any single study on a single course in this domain.

There are only two major methods by which to examine class-level characteristics empirically. Arbaugh (2005) uses one. The other, meta-analysis, takes advantage of the efforts of previous researchers by summarizing observed effect sizes across studies and correcting for sampling error. This way, study-level characteristics can be treated as moderating variables. Studies can be classified based upon their attributes (e.g. “Did the same instructor teach both the in-person and online course?”)

and the true sources of variance in the overall differences observed can be broken down. Only one such effort has been made in personnel psychology (Sitzmann, Kraiger, Stewart, & Wisher, 2006), but several aspects of this meta-analysis make its results difficult to interpret.

Critique of Sitzmann et al. (2006)

Sitzmann et al. (2006) presented the first meta-analysis of differences in outcomes between traditional and web-based instruction in I/O psychology, with a goal of "generalization to a population of adults participating in workplace training" (p. 634). They collected data from 168 courses in 96 studies between 1996 and 2005, including comparisons of traditional vs. fully online courses as well as traditional vs. hybrid courses. Outcomes of interest included post-training declarative and procedural knowledge tests as well as reactions to training. They concluded from this study that web-based instruction was 6% more effective than traditional instruction when examining declarative knowledge (computed from a meta-analytic $d = 0.15$) and equally effective when examining procedural knowledge. Hybrid instruction, on the other hand, was universally superior to traditional instruction in terms of mean knowledge test scores post-training; in no study analyzed did a course comparing traditional and hybrid outcomes produce superior mean outcomes for the traditional course.

They also tested several moderators, the most important of which for this discussion were "experimental design" and "similarity of instructional methods." For experimental design, the authors recorded from each study whether or not random assignment to condition was used. When the study was experimental, they found a substantial difference (*d* = -.26; negative number indicates traditional instruction was superior), while when it was quasi-experimental, they found a substantial difference in the other direction (*d* =.18). For instructional method, the authors recorded for each study if instructional methods were designed to be similar across media ("same"), if no such effort was made ("different"), or if a specific effort was made to make either media superior to the other (also "different"). When instructional methods were similar, they found almost no difference between traditional and web-based instruction (*d* = .04), while when they were different, they found a substantial effect (*d* = .29). From these two findings, they concluded that "our results support Clark's position that media effects in single study research are largely spurious" (Sitzmann, Kraiger, Stewart, & Wisher, 2006, p. 647).

Unfortunately, as stated earlier, many conceptual and technical ambiguities make this conclusion premature and also raise several concerns that will be addressed by the present thesis. I will cover each of these in turn, including research questions addressed in this thesis as I do so.

Generalization to Training. Although Sitzmann et al. (2006) state that this meta-analysis was generated with the goal of generalization to training in employment settings, given the history of research in this area, the vast majority of this literature must have been generated in education technology. This suggests that the literature is predominantly educational in nature and that the major of participants in these studies are students. Indeed, Sitzmann et al. report that 85% of their database is made up of undergraduate and graduate students. Only 15% are actually employees. And additionally, status as a student or employee correlates -0.07 to -0.46 with the other moderators. This does not necessarily threaten the external validity of this meta-analysis to employment settings, but it does raise some question as to the nature of knowledge being trained in these courses and its effects on the results. As very few examples are given of the specific topics being taught, it is difficult to tell just what kinds of subject matter are actually included in these analyses. To be confident that these findings generalize to real work settings, further elucidation is required. Thus, the first research question (RQ1) addressed by this thesis is: "Do medium effects vary as a function of study setting (students vs. workers as trainees)?"

Fixed- vs. Random-Effects Meta-Analysis. A major concern when conducting meta-analysis is the choice of a fixed-effects or random-effects meta-analytic method. Fixed-effects meta-analysis makes an assumption that the population effect sizes (δ, in this case) underlying observed effects are homogenous. For example, in Sitzmann et al.'s (2006) meta-analysis of overall effect size, the use of a fixed-effects model would reflect an assumption that all observed variance amongst observed effects surrounded a single "true" difference between online and traditional instruction. In almost all research, this is an unwise assumption. Confidence intervals computed from fixed-effects meta-analysis are often substantially narrower than they should be, ultimately leading to an overstatement of the precision of observed results (Hunter & Schmidt, 2000).

Regardless, Sitzmann et al. have several specific hypotheses regarding moderators, and thus a random-effects model is clearly appropriate. They utilize the meta-analytic methods of Hedges and Olkin (1985), which include both random-effects and fixed-effects models, although their random-effects models are rarely used (Hunter & Schmidt, 2004). Sitzmann et al. do not report which model they use. This is certainly not conclusive, but it does raise some concern about the precision of these findings worth further investigation. Thus, the second research question (RQ2) addressed by this thesis is: "Will a random-effects meta-analysis produce estimates of similar precision to those estimates obtained by Sitzmann et al.?"

Online vs. Hybrid Courses. Sitzmann et al. (2006) draw several lines when cutting down their meta-analytic database to their final dataset of interest. First, comparisons of traditional and hybrid courses are separated from comparisons of traditional and fully online courses. Next, comparisons examining declarative knowledge are separated from those examining procedural knowledge. After this culling, all moderator analyses are conducted on declarative knowledge outcomes in traditional vs. fully online courses only. This makes the specific rationale behind categorization as a hybrid course very important to the specific composition of their meta-analytic database.

No specific criteria are given as to what Sitzmann et al. consider a hybrid course. As established earlier, most of the literature included here comes from education technology and takes place in a college course. As such, web-supported classes (1-29% online, often through the use of course management software like D2L or WebVista) are likely quite common and would furthermore be quite different from hybrid/blended courses (30-79% online). By Sitzmann et al.'s coding, were web-supported classes considered online or traditional? Were hybrid courses with a minor in-person element considered fully online courses, as they would be by the Sloan Consortium's definitions (Allen & Seaman, 2006)? These questions are left unaddressed. Because of this, it is difficult to know precisely what kind of courses fall into these categorical labels, and thus, on which kinds of courses these moderator analyses were actually conducted.

As much as this literature occurs in educational settings, it is expected that many courses are what Allen and Seaman (2006) would consider "web-facilitated," where the majority of the course takes place in-person with a small proportion of material delivered online. In these courses, material is not delivered predominantly through the web, and is closer in design to a traditional course than a web-based course. A more useful way to examine this distinction, then, would be through determining the specific proportion of the course taking place online and examining its effect. If the specific proportion cannot be determined, then categorization into the Allen and Seaman (2006) framework would be a more useful split than that of Sitzmann et al. (2006). Thus, the third research question (RQ3) addressed by this thesis is: "What effect does the degree to which a course is online have on its effectiveness?"

Declarative vs. Procedural Knowledge. The second cut in Sitzmann et al.'s (2006) meta-analytic database was made between declarative versus procedural knowledge outcomes. They define declarative knowledge as "trainees' memory of the facts and principles taught in training and the relationship among knowledge elements…[while] procedural knowledge refers to information about how to perform a task or action" (p. 627). This is an oversimplification. They take these descriptions from Kraiger, Ford and Salas (1993), who used cognitive psychology to inform new theories about training evaluation. Unfortunately, this paper has been misinterpreted; splitting training outcomes by these two types is unwise. To demonstrate this, I must first delve a bit into cognitive psychology.

When industrial psychologists use the term "knowledge," they are more specifically referring to information stored in long-term memory. Long-term memory itself is split into two general categories with a few familiar labels: declarative memory (consisting of episodic and semantic memory) and implicit memory (consisting of repetition priming and procedural memory). N. J. Cohen and Squire (1980) were the first to adopt the term "declarative" to describe "information that is data-based" (p. 207). This type of information is accessed whenever a person is making a conscious recollection of facts previously learned (semantic memory) or events experienced (episodic memory). Declarative memory is accessed differently than implicit memory, which is a subconscious process. A common example of implicit memory is reading as a native speaker of a language. Such readers do not, moment-to-moment, consciously access rules about grammar, spelling, and punctuation, and yet they are still able to engage in reading using those rules. Most of this information was initially taught to the reader-in-training as discrete facts; in other words, declarative memory was built through conscious learning of facts, but then that knowledge was proceduralized into implicit memory through practice. Thus, the specific subtype of implicit knowledge this refers to is called procedural memory.

Neurological evidence for this distinction among memories is strongly supported by the study of people with anterograde amnesia (N. J. Cohen & Squire, 1980), which is defined by an inability to create new declarative (long-term) memories. For these people, only working memory is available as a basis for forming procedural memory. In these studies, researchers can teach these people new basic skills, including perceptual-motor skills, mirror drawing, and maze navigation, but the learner will have no memory of the learning itself, despite the fact that their procedural memory (and ultimately, skill at the task) will improve with practice. Acquisition of declarative knowledge must occur in order for that knowledge to be proceduralized (Ackerman, 1987), so teaching them complex tasks is much more difficult. With only working memory available, people with anterograde amnesia can only hold relevant instructions in their minds for 15 to 30 seconds at a time, preventing the formation of a base of declarative knowledge (stored in declarative memory) from which complex skills could develop.

Declarative and procedural knowledge, as used by Sitzmann et al. (2006), is a substantial oversimplification of this substantial literature in cognitive psychology. It is true at a very basic level that “declarative knowledge refers to trainees’ memory of the facts…[and] procedural knowledge refers to information about how to perform a task” (p. 627). But it is unwise to categorize the outcomes of entire courses based upon this framework. Sitzmann et al. specifically code procedural knowledge outcomes as “the ability to perform the skills taught in training” (p. 635), but there is virtually no practical training outcome where nothing but procedural knowledge would be taught. For example, in one study included in Sitzmann et al.’s meta-analytic database, a traditional and web-based version of an undergraduate statistics course are compared by examining differences in final course grades (Arvan, Ory, Bullock, Burnaska, & Hanson, 1998). Undoubtedly, this statistics course included and evaluated elements of both declarative knowledge (e.g. “what is a statistic?,” “what are the steps involved in a t-test?”) and procedural knowledge (e.g. “given a dataset, how do I analyze it?”). No information is given in the paper to help determine into which category such a study would ultimately fall. Sitzmann et al. furthermore include any “written test that required trainees to demonstrate memory of the steps required to complete the skills taught in training” (p. 635) as a procedural knowledge outcome, but this is clearly a recitation of declarative knowledge. Given these points, it is unclear as to what specific outcomes are actually included in which category in this meta-analysis. It is furthermore unclear if the specific operationalization reflects a real distinction among training outcomes and draws serious question as to the interpretability of these findings.

Given this, it is still important to break down outcomes into specific categories of capabilities to be trained. An important component of the needs assessment process is to determine the type of capability to be trained and design training to maximize outcomes based upon that assessment. If capabilities taught in training truly differ by category, and if types of training are differentially effective across categories, the distinction should be made solidly. Campbell and Kuncel (2001) present a more useful framework for practical training outcomes than the declarative/procedural knowledge distinction. Capabilities are split into four categories. First, knowledge in general might be trained, including knowledge of labels, facts, rules, procedures, and other discrete pieces of knowledge. Second, observable skills might be trained, including cognitive, psychomotor, physical, interpersonal skills, and other applications of knowledge that accomplish specific goals or solve specific problems. Third, problem solving skills might be trained, including meta-cognition, the use of heuristics, or any other application of knowledge or observable skills to an ambiguous problem supported strategies not specific to the problem. Fourth, any attitude or belief might be trained, such as self-efficacy or racial attitudes. Thus, the fourth research question (RQ4) addressed by this thesis is: "Do online courses differ in effectiveness from traditional courses, and if so, do these differences change based upon the type of capability trained?"

Experimental Design and Self-Selection. Perhaps the most troubling aspect of Sitzmann et al.'s (2006) meta-analysis is the treatment of experimental design. They defined quasi-experimentation by stating "research reports utilized a quasi-experimental design when trainees self-selected into WBI or CI" (p. 636), but this is not the definition of quasi-experimentation. Quasi-experimentation only necessarily involves nonrandom assignment to conditions, which includes self-selection into conditions, but may include other methods of nonrandom assignment as well. For example, in one study included in Sitzmann et al.'s meta-analytic database, the traditional class being compared was held in a local community college while the web-based class was held in other colleges specifically structured around distance learning (Barker, 2002). This study is clearly quasi-experimental in nature, as learners were not randomly assigned to conditions, and yet these learners did not have any choice as to which medium they would be using. Similar situations occur when courses are switched year-to-year. For example, a course might only be offered in one year as an in-person lecture-based course, and then converted outright the following year to an online course. When learners enrolled in either set of these courses, they would have no choice as to which medium they were experiencing. In other studies, if they were indeed given the choice to enroll in the medium they believed would best support their learning, they might logically experience most positive outcomes. Thus, if Sitzmann et al. (2006) stayed true to their given definition and ignored quasi-experimental designs other than by self-selection, it makes interpretation of the quasi-experimentation moderator impossible, as it confounds nonrandom assignment with willful self-selection into condition.

Furthermore, most conclusions made by Sitzmann et al. (2006), such as those regarding the Clark-Kozma debate, are based upon other moderator results collapsed across the quasi-experimentation moderator. 85% of studies where design was discernable utilized quasi-experimentation, with a substantial difference between findings from experiments and quasi-experiments (d_{qe} = -.26, k = 11; d_{ex} = .18, k = 60). Yet this substantial difference is ignored when examining the instructional methods moderator discussed earlier (d_{same} = .04, k = 16; d_{ex} = .29, k = 37). Essentially, by ignoring any pre-training test information, and in the absence of a control group, they are treating every study as a "posttest-only design with nonequivalent groups," which Cook, Campbell and Peracchio (1990) go so far as to call "generally uninterpretable" (p. 517). Without pre-test information, a control group, or random assignment, all observed differences are attributed to the effect of medium although they may not be caused by medium; there is no way to determine the true cause of these observed differences using this information alone.

Self-selection itself is a useful concern; individuals who self-select into one medium might do so in order to put themselves in a position to best learn the material. Separating self-selection from quasi-experimentation would allow us to examine both aspects of this issue fully. Thus, the fifth research question (RQ5) addressed by this thesis is: "Does self-selection affect the observed differences between traditional and online courses, and do individuals that self-select into online courses systematically differ from those that self-select into traditional courses?" The sixth and final research question (RQ6) addressed by this thesis is: "How do within-person outcomes (increases in knowledge over time within a training medium) inform our interpretation of between-person outcomes (differences in outcomes between media)?"

Rationale for the Present Thesis

Given these difficulties in the interpretation of Sitzmann et al. (2006), the purpose of this thesis is to expand upon Sitzmann et al.'s meta-analytic database and improve upon it, addressing the six research questions raised by a critique of Sitzmann et al. Each of these research questions will be directly addressed by using improved meta-analytic method and coding strategies while simultaneously increasing the reference literature for this meta-analysis with those articles published since Sitzmann et al. By doing this, I am producing more accurate meta-analytic estimates of the true differences between traditional and web-based instruction with clear explanations and rationale as to specific moderators to be investigated.

Method

Literature Search

As a basis for the database to be used in this study, Sitzmann et al. (2006)'s original database was collected and then expanded upon using an identical search strategy. To do this, PsycINFO and ERIC were searched for articles published 2005 – 2009 using the following search string on keywords:

(web or online or internet) and (evaluate or learn or transfer or behavior or performance or knowledge or satisfaction or dissatisfaction or reaction or achieve or outcome)

This resulted in an initial list of 6823 sources from PsycINFO and roughly 1100 from ERIC, which was then reduced to a more relevant list by a review of titles and abstracts for articles likely to meet the inclusion criteria stated below. This decreased the list of potentially relevant studies to 245 total sources, including the original Sitzmann et al. (2006) database. Ultimate review of the articles themselves further reduced this number to 116 sources. In addition to this general search, the reference lists from review articles found during this process were scanned for potential inclusions, a new search using the search term “(elearning or e-learning or web-based)” was run on all journals contributing more than 1 article to the database up to that point (21 sources), and all articles published in the *Journal of Asynchronous Learning Networks* from 2005 to 2009 were scanned by hand. This manual process produced a list of 27 potential inclusions and ultimately added 8 sources.

Twelve articles included by Sitzmann et al. (2006) were not included in the database by this stage, typically because insufficient information was available in the article as published. Each author was contacted for additional information needed to calculate a *d*-value, ultimately adding two articles back to the database.

Thus, this entire process produced a final database containing 126 sources.

Meta-Analytic Database

Inclusion Criteria. As with Sitzmann et al. (2006), all research reports where students or employees were being trained in both traditional and web-based or web-supported training were included. Sitzmann et al. further specified this by stating all reports were included where trainees were participating in training "to prepare them for current or future employment opportunities" (p. 634); however, the specific implications of this statement were never explained, nor were any examples given, so this restriction was ignored.

Studies needed to meet four criteria to be included, similar to those of Sitzmann et al.: 1) the study must compare outcomes between traditional and some type of online training, 2) the article must be available in English, 3) the study must report results in such a way that a *d* statistic can be calculated, and 4) the participants in the study must be nondisabled adults aged 18 or older. Sitzmann et al. (2006) did include a fifth criterion ("training was conducted on a topic that provided job-related knowledge or skills", p. 634), but since this was not operationalized, and considering training topics like self-motivational techniques (Stadtlander, 1998) and students in a lab soldering a circuit board (Alzafiri, 2001) were included, this restriction was re-interpreted as grounds for exclusion: if a study clearly had absolutely no relevance to the workplace, it was excluded.

Independence of Observations. To ensure that the assumption of independence of observation amongst meta-analytic cases is not violated, several specific decision rules were used. First, independent samples contained within the same article were treated as independent cases. Second, multiple outcomes that differed by criterion type (i.e. knowledge, observable skills, problem solving skills, attitudes) were not considered independent. Third, when multiple outcomes did not differ by criterion type, *d* was computed for each effect, and a single mean *d* was included (for example, if a multiple choice knowledge test and written knowledge test were reported separately, d_{MC} and d_{WT} would have been determined for each and averaged) unless a meaningful composite was available, in which case said composite was included alone (for example, if three knowledge tests and two skills tests were reported along with a knowledge composite and skill composite of those tests, the two composites alone would be included in their respective meta-analyses, while a sample-weighted average of the composites would be used in an overall meta-analysis). Fifth, when multiple comparisons were possible from a single source (e.g. if two web-based courses were compared with a single traditional course), then all possible pairs of comparisons were computed and a final sample-size weighted mean *d* was included.

It is important to note that the third rule above deviates substantially from that of Sitzmann et al. (2006). They state that "whenever a single study reported multiple effect sizes based on the same sample for a single criterion, the effect size that was most similar to the other assessments of that particular relationship was used" (pp. 634-635). This seems to indicate that when multiple effect sizes were available, Sitzmann et al. (2006) chose whichever effect size was closest to the mean of other effect sizes in that particular meta-analysis, introducing artificial homogeneity of observed effects. The present approach avoided this confound.

Meta-Analytic Method. The method used to compute each meta-analysis was Hunter and Schmidt's (2004) psychometric meta-analysis. As the comparison of traditional and web-based training is a natural dichotomy, *d* statistics were computed for each comparison by subtracting the mean outcomes of each traditional class from the mean outcomes from any online course and dividing by the pooled standard deviation. Reliabilities of each criterion were recorded where available, but relatively few were reported, and were not used in subsequent analyses. Thus, the only correction applied was for sampling error in the standard deviation of *d*. Meta-analytic results involving pre-test/post-test scores were separated into two sections. First, the post-test scores and pre-test scores between each pair of methods were examined using between-persons *d*-values meta-analyses (e.g. one meta-analysis contained web-based post-test scores compared with traditional post-test scores; another contained web-based pre-test scores compared with traditional pre-test scores). Second, the pre- and post-test scores within each method were examined using a within-persons *d*-values meta-analysis.

Because the majority of the research literature analyzed involved non-random assignment (i.e. quasi-experimentation), there are three primary methods by which to analyze the initial meta-analytic results taking pre-training differences into account. First, as done by Sitzmann et al. (2006), pre-training differences can simply be ignored, and the post-training d alone is the outcome of interest. Second, Δd_B is defined as the difference between two between-subject d's. For example, if the difference between post-training traditional and web-based training outcomes was $d = .20$, and the pre-training difference was .05, Δd_B would be $.20 - .05 = .15$. Thus, this would be interpreted as: web-based training produces outcomes .15 standard deviations higher than traditional training. Third, Δd_W is defined as the difference between two within-subject d's. For example, if the difference between pre-training and post-training outcomes for web-based training outcomes was $d = 2.00$ and the difference between pre-training and post-training outcomes for traditional training outcomes was $d = 1.85$, Δd_W would be defined as $2.00 – 1.85 = 0.15$. Thus, this would also be interpreted as: web-based training produces outcomes .15 standard deviations higher than traditional training.

Meta-Analytic Coding

To ensure consistency in the values assigned to each entry in the meta-analytic database, one person recorded data from each article to be included while a second coded 20% of those articles, randomly selected from the full list. Percentage agreement and Cohen's κ statistics for each appear in Table 1. Percentage agreement is a liberal estimate of agreement while Cohen's κ is a conservative estimate, so actual inter-rater reliability is assumed to be between the two values.

Training Type. In order to ensure transparency of the skills being taught in training, to assess generalization of these results to a work setting, a simple table of all topics trained is found in Table 2, categorized based upon trainee type, student or employee. Study setting, as lab or field, was also recorded for each.

Course Hybridization. To accurately assess the degree to which a course is online, it would be preferable if a specific “percentage online” could be determined for each course. This degree of detail was not present in these articles, however, so as a practical concession, a slightly modified version of the Allen and Seaman (2006) framework was used. Courses were categorized into one of four categories. First, courses were considered “traditional” if there was no online component whatsoever. Second, courses were considered “web-facilitated” if online material was used only to support learning in the traditional setting. Third, courses were considered “hybrid” if material that was meant to provide instruction was presented both in a traditional setting and online. Fourth, courses were considered “fully online” if there was no in-person component whatsoever. By these definitions, a course using Moodle to make lecture notes and the syllabus available online would have been considered a “web-facilitiated” course, while a course using Moodle to deliver lecture videos in addition to in-person lecture content would have been considered a “hybrid” course.

Because hybridization can also have an effect on experimental design, the type of comparison regarding hybridization will also be recorded as “complete” or “addition.” For example, Alzafiri (2001) compared a traditional and hybrid course, but the hybrid course was actually identical in content to the traditional course, with web-based content added *in addition* to the traditional content. This was not a pure comparison of subjects entering two distinct courses; instead, this was a quasi-experiment conducted with a control group, where the entire manipulation was done in addition to what occurred already in the control condition.

Learning Outcomes. Categorization of learning outcomes into knowledge, observable skills, problem solving skills, and attitudes were made based upon Campbell and Kuncel's (2001) framework. Knowledge was defined as any discrete pieces of information (e.g. labels, facts, rules, procedures, plans, goals). Observable skills were defined as the use of knowledge to solve specific problems or achieve specific goals. Problem solving skills were defined as the use of knowledge and observable skills to solve ambiguous problems with no clear path to a solution. Attitudes were defined as any belief or personality-based characteristics (e.g. self-efficacy, racial attitudes, sexist beliefs). A fifth outcome, perceptions, was also added to reflect measures where the learner self-reported learning outcomes.

This coding scheme directly addresses concerns with the use of declarative and procedural knowledge by Sitzmann et al. (2006). To contrast the present method with the previous method, consider that in one paper, Sitzmann et al. coded a "declarative knowledge assessment [which] consisted of a multiple-choice and fill-in-the-bank examination designed to assess understanding of the concepts taught…the procedural knowledge assessment required trainees to perform the software application skills taught in training" (p. 635). Undoubtedly, the "procedural knowledge" outcome described here involved declarative knowledge as well; trainees needed to know what the program was, what functions were available, etc. in order to use that program. Under the present framework, the multiple-choice/fill-in-the-blank test and computer-usage test would have been defined more clearly as knowledge and skill tests, respectively.

Reactions. To more fully replicate and expand upon the efforts of Sitzmann et al. (2006), reactions to training will also be recorded and categorized by type of reaction measure, if possible. It was expected that more reliabilities would be available here, and more were found; however, this number was still too low (12%) to reasonably produce an artifact distribution, preventing corrections for attenuation due to unreliability in addition to sampling error. Aggregation across reaction measures occurred in a similar fashion as with learning outcomes.

Experimental Design. To fully explore the impact of experimental design in regards to self-selection, the use of random assignment (experimental vs. quasi-experimental) and the ability for trainees to self-select into condition knowingly (self-select vs. experimenter-select) was recorded. It is important to note that while quasi-experimentation is necessary for self-selection, self-selection is not necessary for quasi-experimentation, so these two variables cannot be fully crossed. Thus, the final categorization scheme for experimental design is experimental vs. quasi-experimental with self-selection vs. quasi-experimental without self-selection. Self-selection was further categorized by its own type (course vs. program); in undergraduate programs, some students choose to enroll in specific online courses while others enroll in entirely online programs. However, only one study found examined program differences, so this subcategory was ultimately dropped.

Because the ultimate effect of self-selection would be to alter the knowledge or skill level of participants entering training, differences in pre-training scores were also recorded. Only parallel or identical measures were accepted as pre-test scores; variables used as covariates were not recorded. To attempt to explain any variation in pre-test scores, any and all differences in individual differences (e.g. personality, cognitive ability) between those self-selecting into conditions were also recorded.

Other Variables. Several other minor moderators were included. The year that the course was designed was coded as an indicator of the age of the online program. If the year that the course was designed was not available, the year that the online course was administered to learners was used instead. E-Location was also coded, which entailed whether learners were given their choice of online training location (e.g. at the library, at home) or whether it was restricted in some way (e.g. trainees can take the course at any time, but must do so in an approved computer lab). Finally, two course design moderators were recorded: the use of learner control and human interaction in the fully online courses were recorded as binaries (yes or no).

Results

Overall Descriptive Meta-Analyses with Design Moderators

A full intercorrelation matrix of all study variables appears in Table 3.

To address the various research questions, several meta-analytic summary tables were produced. First, Table 4 presents the overall between-persons differences in post-tests and pre-tests between traditional courses and each online course type (web-facilitated, hybrid, and fully online), finding substantial differences between types. Table 5 shows both between-person and within-person differences for traditional vs. online courses. Most notable here is the difference between the two sets of findings. By the between-person results alone, fully online instruction appears as effective as traditional instruction (d_{Post} = .17; d_{Pre} = .17, $\Delta \underline{d_B}$ = .00). When examining the within-person effect sizes, however, online instruction appears less effective (d_{T} = 2.19; d_{FO} = 1.99; Δd_W = -.20). To try to explain these differences, post-training between-person differences were further split by the presence of pre-training between-person differences; in other words, the last four lines of this table all refer to the same twenty studies. Looking only at the studies reporting pre-training differences, it once again appears that online training is less effective, although the difference is smaller (d_{Post} = .12; d_{Pre} = .17, Δd_B = -.05). This implies that there may be something else substantively different about studies that include pre-training differences from those that do not.

To more fully investigate the effect of the large pre-training difference found for traditional vs. online courses, Table 6 separates pre-training differences by experimental design. It would not be expected that experimental designs and quasi-experimental designs without self-selection would show mean pre-test differences across studies, and this is confirmed here. Only quasi-experimentation with self-selection shows a difference in pre-training scores (d = .19).

To investigate the effect of trainee type, Table 7 presents differences in pre- and post-training scores split by trainee type: student or employee. Very few organizational studies are available in this comparison (k = 11), and substantial differences in post-training d-values bring some cause for concern. Organizational studies show fully online courses to produce substantially inferior outcomes to traditional courses (d_{Pre} = .11; d_{Post} = -.17; Δd_B = -.28), while educational studies show little difference (d_{Pre} = .17; d_{Post} = .19; Δd_B =.02). These differences alone are difficult to interpret, however, as pre-training differences differ by experimental design, which is in turn correlated with trainee type (the correlation between the use of experimentation and trainee type = .31; Table 3).

Table 8 presents differences split over lab and field. Although the number of studies conducted in lab settings (k = 2) are far too few to interpret on its own, the differing direction of this moderator in comparison to field studies is worth future investigation.

Table 9 replicates Table 5 but using hybrid courses. The findings here are reversed from Table 5; hybrid courses appear superior in both between-persons (d_{Post} = .31; d_{Pre} = .04; Δd_B = .27) and within-persons comparisons (d_{H} = 1.41; d_{T} = 1.28; Δd_W = .13), although k's are much smaller. Post-training scores also differ between those studies that report pre-training scores and those that do not. Table 10 replicates this again but with web-facilitated courses. The reversal of findings across between- and within-person results is present here as well; web-facilitated courses appear superior by the between-person analyses (d_{Post} = .14; d_{Pre} = -.03; Δd_B = .17), but slightly inferior by the within-person analyses (d_{WF} = .85; d_{T} = .91; Δd_W = -.06).

Table 11 examines within-person effects across training types, comparing students and employees as trainees. The observed differences are opposite in pattern to Table 7; fully online instruction appears superior for employees (d_{FO} = 1.61; d_{T} = .62; Δd_W = .99) while traditional instruction appears superior for students (d_{FO} = 2.00; d_{T} = 2.27; Δd_W =.27). Unfortunately, the employee estimate is unlikely to be stable considering the small *k* for organizational studies (*k* = 2).

Descriptive Meta-Analyses Split by Criterion

To investigate differences between criteria (knowledge, observable skills, problem solving skills, attitudes, and perceptions), Table 12 examines knowledge criteria in traditional versus online courses in the same fashion as Table 4; fully online courses appear superior by the between-person analyses (d_{Post} = .27; d_{Pre} = .14; Δd_B = .13), but equally inferior by the within-person analyses (d_{FO} = 1.88; d_{T} = 2.04; Δd_W = -.16). Table 13 gives the same treatment to hybrid courses, while Table 14 does so with web-facilitated courses. Table 12 has the largest numbers of studies of the set, and as the moderator analyses become increasingly specific, so do *k*'s grow alarmingly small. In Table 15, observable skills can only be examined at the post-training level; insufficient information is available pre-training (*k* = 0 or *k* = 1 in other cells). Post-training scores vary widely, but small *k*'s and lack of pre-training differences make them difficult to interpret. In Tables 16 and 17, post-training and pre-training differences are examined for problem-solving skill criteria, for traditional vs. fully online and traditional vs. web-facilitated courses, respectively, but low *k* again makes interpretation difficult. Only one study reported problem solving skill criteria for traditional vs. hybrid courses, so this study was not included. Attitudes as an outcome were only reported in two studies; all findings regarding attitudes are found in Table 18.

The perceptions criterion is examined in Table 19. This criterion is of particular interest because it is not a capability to be trained as specified in the Campbell and Kuncel (2001) model and was discovered unexpectedly during coding as the sole outcome of interest for several studies included by Sitzmann et al. (2006). It consists solely of trainee's opinions regarding their understanding of the material. For example, if a trainee was asked to self-report their understanding of the subject matter on a 10-point scale before and after training, such outcomes would be classified as perceptions. Here, the within-person differences appear to reveal no benefit of either medium (d_{FO} = 3.82; d_{T} = 3.88; Δd_W = -.06), while the between-person differences reveal a disadvantage for online courses (d_{Post} = -.16; d_{Pre} = .09; Δd_B = -.25). If this is primarily based upon quasi-experiments with self-selection, as would be guessed, this suggests that trainees entering online courses believe they know more about the topic already than do their traditional counterparts, but ultimately believe they learned less of the material.

In Table 20, the focus switches to differences in reactions to training, which are typically measured by course satisfaction measures or college end-of-semester student evaluation forms. Here, it appears that differences across the various experimental designs between traditional and fully online courses are quite wide, although all are negative. On average, people have poorer reactions to fully online courses than traditional ones (d = -.21). In Table 21, a similar approach is taken on comparisons of traditional and web-facilitated courses, although the very low number of studies (k = 6) examined makes interpretation difficult.

Individual Differences Underlying Pre-Training Capability Differences

In order to determine the underlying source of variation in pre-test scores for quasi-experiments with self-selection, these studies were re-coded for any pre-training individual differences. Age and gender were both included, as well as a variety of psychological variables. Because the psychological scales varied greatly in content, they were qualitatively sorted into categories in order to increase k for each examination. The list of scales and their categories appear in Table 22. Meta-analytic differences in pre-test scores between traditional and fully online courses on these individual differences appear in Table 23, while differences between traditional and hybrid courses appear in Table 24. k is low for all comparisons, again making interpretation difficult. However, a sizable difference in age appears between traditional and fully online courses (d = 1.61), confirming many anecdotal observations made in educational technology that online students tend to be non-traditional, older students.

Rationale for Adjustment to Post-Training Scores by Pre-Training Differences

Considering the large observed average difference in pre-test scores between those self-selecting into online and traditional instruction (*d* = 0.19; Table 6), as well as the differences between studies reporting pre-training scores versus those that do not (Table 5), disentangling the effect of pre-tests is vital to understanding most of the moderators of interest. Because such a small number of studies with pre-test scores were available (*k* = 29 across online course types), limiting the meta-analysis to only these studies is exceptionally restrictive. To retain the advantages of a large dataset while addressing pre-test differences, Wortman (1992) suggests adjusting post-test effects by pre-test effects. In Table 6, differences of .02, .01, and .19 were found between traditional and fully online course pre-training scores for experiments, quasi-experiments without self-selection, and quasi-experiments with self-selection, respectively. Thus, all analyses from this point forward (all results in Tables 25 to 31) consider only comparisons of traditional and fully online courses post-tests, for which the post-test scores from quasi-experiments with self-selection have been adjusted down by .19 standard deviations to account for average pre-training differences in outcomes. Wortman (1992) indicates that reliability corrections should be applied before post-test effects are adjusted for pre-test effects; otherwise, post-test differences may be slightly over-corrected. Thus, as reliabilities are unavailable in these meta-analyses, this may represent a slight over-correction.

Because of this adjustment, Tables 25 to 31 are no longer entirely descriptive. Instead, there are two assumptions. First, it is assumed that pre-training differences are uncorrelated with the other moderators of interest. Generally, correlations with the use of this experimental design are small (again, see Table 3), but this is not definitive. Second, it is assumed that had those pre-training differences not existed, post-training differences would have been reduced by an equal amount. In a sense, this assumes perfectly normal distributions of outcomes (that no ceiling effects occurred in the measurement of training outcomes), an assumption for which there is no way to confirm. Thus, all analyses from here should be interpreted with these caveats in mind.

Meta-Analyses of Adjusted Post-Training Scores and Moderators

Table 25 contains post-test differences across the experimental design moderator. The difference in post-tests (d = .08) reflects a slight overall advantage for fully online courses over traditional courses. Findings from experimental studies and quasi-experimental studies without self-selection would be unaffected by the d = .19 downward adjustment, so these moderators reflect actual post-test differences among modes. The quasi-experimental with self-selection moderator reflects the adjustment; it appears from this that traditional courses are slightly more effective than fully online courses when the design is quasi-experimental with self-selection.

Table 26 contains the setting and trainee type moderators examined before. Lab vs. field appears similar; the difference between students and employees as trainees is still quite large.

Year of publication and the year that the online course was designed are both treated as moderators in Table 27. Year that the online course was designed is meant to reflect the modernism (and perhaps sophistication) of the online training program, however it is much less commonly available than the year of publication (or production, if unpublished), so both are included here. In general, it appears that older courses show smaller (or negative) differences, and newer courses show larger differences between traditional and fully online courses. However, the negative *d* for 2006-2009 suggests this may not be a linear relationship. It can be safely concluded from this table that the year makes a difference; the precise nature of that difference is unclear.

Student type is included as a moderator in Table 28, which might be considered a proxy for job complexity. Undergraduates show a positive difference between courses modalities, while graduates show a negative difference, implying that more complex courses may be served more poorly in an online environment. Scope is also examined in this table; studies on students are sometimes semester-long courses, and sometimes individual (1-day, for example) interventions. The difference between these is not as large as expected.

In Table 29, outcomes are split by criterion type. According to this table, knowledge and attitudes are trained equally well by modality, observable skills are trained less well online, and problem solving skills are trained better online.

Differences in publication status appear in Table 30. Unsurprisingly, published studies tend to show larger differences than unpublished ones. The present database is over 25% unpublished, however. E-Location is another moderator of interest in this table; studies in which the learner is able to complete the training at any location show larger differences than those in which the learner is restricted. This is difficult to interpret, however, considering the large correlations between e-location and other moderators (see Table 21).

Finally, two moderators included in Sitzmann et al.'s (2006) examination are included here, although they should be interpreted with some caution due to low inter-rater agreement (see Table 1) and high inter-correlations with other moderators (see Table 3). The learner control moderator here reflects a precisely opposite conclusion from Sitzmann et al. (2006). The present results suggest that the use of learner control is detrimental for web-based learning outcomes. The results from the human interaction moderator were similar, however. Generally, findings do not differ across the two conditions.

Discussion

RQ1: Differences Due to Study Setting

Research Question 1 is, "Do medium effects vary as a function of study setting (students vs. workers as trainees)?"

In general, they do, and the differences are a cause for concern. In Table 7, without the d = .19 downward correction, media differences where employees are trainees (Δd_B = -.28) are much weaker that media differences where students are trainees (Δd_B = .02). This implies that in terms of final outcomes, students benefit from being in online courses while employees benefit from being in traditional courses. However, when examining within-person differences in Table 11, it appears that the online courses are strongly preferred (d_{FO} = 1.61; d_T = .62; Δd_W = .99) for employees. Here, students gain more in traditional settings, while employees gain more in online settings, although this must be interpreted with some caution due to the small number of organizational studies providing pre-training scores (*k* = 2).

Post-test differences with the correction applied add another piece of information to this puzzle. Table 26 provides these values, and they are quite similar in pattern to those presented in Table 7. Generally, employees tend to perform better on outcome measures after a traditional course while students tend to perform better after a fully online course. Considering the small organizational *k* in Table 11, the results from the meta-analyses using adjusted post-test scores are probably the most reliably interpreted.

The precise mechanism behind these differences is not clear. On one hand, it may simply be another manifestation of the year moderator examined in Table 27. Organizations may adopt new online training practices more slowly than universities, and thus use less modern (and less sophisticated) training packages. Alternately, the freedom provided to college students by using online college courses may be the causal mechanism producing better outcomes for students. This freedom of access may be less valuable to employees, who are completing their training while at work. From this data alone, it is impossible to tell. But at the least, it can be concluded that there are substantial differences between educational and organizational settings. The assumption of the generalizability of educational research in this domain to real organizations may be unwise.

RQ2: Random- versus Fixed-Effects Meta-Analyses

Resarch Question 2 is, “Will a random-effects meta-analysis produce confidence intervals of similar precision to those estimates obtained by Sitzmann et al.?”

Sitzmann et al.’s (2006) estimate of declarative knowledge differences between classroom and web-based instruction shows a confidence interval of $.11 \leq d \leq .19$, with a standard error of .02. Due to differences in the way that moderators were determined, there are two possible comparisons in the present study. The first is the comparison of all studies examining differences between traditional and fully online courses, found in Table 4. The second is this comparison limited to knowledge outcomes alone, found in Table 12. Each of these was handled separately.

Using information from Table 4, the standard error was estimated as .039 ($.37 / \sqrt{88}$), using the computation recommended by Hunter and Schmidt (2004, p. 206). Using information from Table 12, the standard error was estimated as .060 ($.43 / \sqrt{52}$). Because the standard error is strongly affected by k, because k differs between the Sitzmann et al. (2006) estimates (k = 71) and these (k = 88, k = 52), and because the databases of studies are similar but not identical, it is difficult to say conclusively that Sitzmann et al. used a fixed-effects meta-analysis. However, even the more conservative estimate of the standard error (.039) in the present study would produce a confidence interval .15 units wide, compared with the .08 units wide Sitzmann et al. (2006) estimate. Thus, the Sitzmann et al. estimate is at best 53% narrower.

The present study uses only random-effect meta-analyses. Most confidence intervals, if computed, would be quite wide. It should be noted that credibility intervals are reported in the tables; these provide different information from confidence intervals. Generally speaking, in a 95% confidence interval, there is a .95 probability that the interval contains the true mean *d*, while a 95% credibility interval shows the range within which 95% of values are expected to fall, net of sampling error.

RQ3: Differences due to Hybridization

Research Question 3 is, "What effect does the degree to which a course is online have on its effectiveness?"

Examinations of the effect of the degree to which a course is online are difficult in the present study due to small cell sizes in many of the analyses. From Table 4, we might conclude there are relatively small pre-training differences between those choosing hybrid and web-facilitated courses versus those choosing traditional courses. This is logical, as most people choosing online courses due to convenience factors would likely only do so to avoid having to go to a classroom entirely. However, the numbers of studies here are still small (k = 3, k = 6, respectively), so caution is required, especially in interpretation of Tables 9, 10, 13, 14 and 17. First, hybrid courses appear superior to traditional courses both in terms of between-person differences (d_{Post} = .31; d_{Pre} = .04; Δd_B = .27; Table 9) and within-person differences (d_H = 1.41; d_{T} = 1.28; Δd_W = .13; Table

9), with similar values for knowledge-specific outcomes (see Table 13). As hybrid courses represent a meaningful portion of instruction delivered both in-person and online, this might be simply because more information in more formats is available to trainees. They are, in a sense, receiving *more* training, so it makes sense that their outcomes would be stronger. Web-facilitated courses, too, appear superior to traditional courses in terms of between-person differences (d_{Post} = .14; d_{Pre} = -.03; Δd_B = .17). However, when examining within-person differences (d_{WF} = .85; d_{T} = .91; Δd_W = -.06), the relationship again reverses. The same logic applies here; a web-facilitated course is most often a traditional course with materials placed online. So if a student missed a day of class, in a web-facilitated course, they could simply go online and retrieve the lecture notes that they missed. A student in a traditional course would have fewer options to learn that material. It is peculiar that the within-person *d*'s in Table 10 are so small compared to most other within-person *d*'s in this study; there may be some unknown factor at play here suppressing these values.

The differences in within-person *d*-values looking at web-facilitated courses are relatively small (Δd_W = -.06), and the differences in post-test d's are relatively large (Δd_B = .14). From this perspective, it seems that web-facilitated courses might produce superior learning outcomes to traditional courses, although it is not quite clear. But for hybrid courses, it is even less clear. It may be that hybridization improves outcomes, or that simply *more training* is occurring in such courses. Future studies of hybrid courses should be careful to differentiate whether there are qualitative differences between the hybrid course and its comparison group, or whether some portions of the course are simply being converted to an online format.

RQ4: Differences due to Criterion

Research Question 4 is, "Do online courses differ in effectiveness from traditional courses, and if so, do these differences change based upon the type of capability trained?"

In general, outcomes do differ somewhat. When examining differences between traditional and fully online courses for knowledge, the same reversal as when looking at between- and within-person differences occurs. When examining between-person differences, fully online courses appear superior (d_{Post} = .27; d_{Pre} = .14; Δd_B = .13*;* Table 12). When examining within-person differences, traditional courses appear superior (d_{FO} = 1.88; d_{T} = 2.04; Δd_W = -.16; Table 12). When examining observable skills, where it would be expected that online courses would be least effective, only post-test comparisons are available, and there is no difference (d = .00; Table 15). When looking at problem solving skills, online instruction is consistently inferior. When looking at between-person differences, traditional courses appear superior (d_{Post} = .27; d_{Pre} = .40; Δd_B = -.13*;* Table 16). When looking at within-person differences, traditional courses also appear superior (d_{FO} =.79; d_{T} =.83; Δd_W = -.04; Table 12), although there is not much data to examine (k = 2).

It was expected that observable and problem solving skills would be more difficult to train online, and this is confirmed. Tests of observable skills in this meta-analysis are quite varied and often include physical components (e.g. psychomotor tests, mathematical computation, technical skills licensure, graphing). Problem solving skills as measured are more rarely physical (e.g. essay-writing, course projects, reasoning, marketing project), but are of greater complexity and require more human interaction. Thus, while the online environment might make communication with group members more accessible and learning materials more available, it does not appear to ultimately benefit learning. In the training of problem solving skills, process losses in group work might also be more severe than they would be in-person. Unfortunately, from this meta-analysis alone, it is unclear. Further research is needed to differentiate between training outcomes and target training interventions toward them.

Only two studies examined differences in attitudes, so conclusions drawn are limited. However, from these two, it does not appear that there are any differences in course effectiveness (d = .00; Table 18).

Perception criteria are not included in Campbell and Kuncel's (2001) taxonomy of training capabilities because they are not capabilities. It is not the purpose of most training programs to have trainees only think that they learned something. Each perception is likely linked to a real (untested) training outcome. In a sense, perception criteria might be best characterized as some unknown combination of self-efficacy and capabilities gained. In this meta-analysis, trainees tend to believe that they have learned less (perhaps reflecting lower self-efficacy and learning outcomes) by completing fully online training as opposed to traditional training (d = -.16; Table 19).

There are pre-training differences as well, however. Individuals entering online training programs already have higher perceptions than do those entering traditional programs (*d* = .09; Table 19). This implies that trainees entering online training programs believe they know more entering the program and less exiting it than do their traditionally-instructed counterparts. When examining this using the prior analysis scheme, it appears that online courses produce poorer outcomes when examining both between-subject outcomes (d_{Post} = -.16; d_{Pre} = .09; Δd_B = -.25) and within-subject outcomes (d_{FO} = 3.82; d_{T} = 3.88; Δd_W = -.06), although the small *k* for within-person studies (*k* = 3) once again limits generalizability.

The findings from the self-selection-adjusted post-training scores (Table 29) generally confirm the above observations. There are no post-training differences between traditional and fully online courses for attitudes outcomes (*d* = -.02; d = .00, respectively). Observable skills slightly favor traditional courses (*d* = -.12), while knowledge and problem solving skills favor fully online courses (*d* = .19 and *d* = .25, respectively). And finally, perceptions of learning are generally lower following fully online courses (*d* = -.20).

Overall, it can definitively be concluded that online courses as they are typically studied are differentially effective in regards to criteria. It cannot be concluded, however, that this is an inherent property of the media. Just as there are differences over time in effectiveness, techniques for properly translating observable skill training into an online environment may simply not yet be available. Future research should concentrate on making the distinction between criteria and identifying techniques to be used to maximize the effectiveness of online training for all types of training capabilities.

RQ5: Differences due to Self-Selection

Research Question 5 is, "Does self-selection affect the observed differences between traditional and online courses, and do individuals that self-select into online courses systematically differ from those that self-select into traditional courses?"

Self-selection has a substantial and noticeable impact on observed differences. According to Table 6, experiments and quasi-experiments without self-selection show virtually no pre-test differences (d = .02, d = .01, respectively), while quasi-experiments with self-selection show substantial differences (d = .19).

Unfortunately, attempts to determine correlates with pre-training individual differences were limited. Few studies reported pre-test scores, and still fewer also reported individual differences split by group. Table 23 provides the few differences found when examining traditional and fully online courses, including most interestingly a sizable difference in the age of participants (d = 1.61). This was expected; anecdotally, older, non-traditional students are expected in online courses in college settings as these courses are more convenient to those working outside of school. Although psychological differences were observed in affect and computer experience, the effect sizes and k's were too small to make any definitive conclusions as to any underlying differences. This makes a correlation matrix between pre-training scores and these outcomes impossible to interpret considering the small cell sizes. Thus, we have no answer to RQ5. Although there are clearly differences of some sort, and these differences likely reflect underlying differences in psychological traits or states, there is insufficient information here to make any conclusions. At the least, future studies using quasi-experimentation in this context should include such information in order to track down the real differences between the populations entering each course.

RQ6: Within-Person Outcomes vs. Between-Person Outcomes

Research Question 6 is, "How do within-person outcomes (increases in knowledge over time within a training medium) inform our interpretation of between-person outcomes (differences in outcomes between media)?"

This is a difficult question considering the sometimes conflicting information provided by these meta-analyses. In Table 5, the post-training and pre-training differences between traditional and fully online courses are identical (d = .17). Alone, this might be considered "no difference" (Δd_B = .00); the groups, on average, differed equally both before and after training. However, when examining the within-person outcomes, the traditional course appears more effective than the fully online course (d_T = 2.19; d_{FO} = 1.99; Δd_W = -.20). There are several reasons this might happen. First, studies that report pre-test scores may vary systematically from those that do not. Second, there may be ceiling effects such that training programs are only designed to reach a certain level of mastery. Third, pooling the pre-training and post-training SDs may create a biased estimate of the overall effect. Fourth and finally, the programs may really be differentially effective.

It is impossible to differentiate between the four given the results of this meta-analysis alone, however it does highlight the importance of making such a distinction in the training literature. In much training research as published, the ultimate goal of the training program is never articulated along these lines. Is the goal to increase knowledge in a particular domain? Is the goal to reach competency or mastery of a particular task, which once mastered, need only be maintained? Given a particular traditional or online training program, does this distinction make a difference? It is logical that if mastery is the goal, post-training outcome differences between traditional and online training programs are irrelevant. Instead, the efficiency (time-to-mastery) and resources used by the two programs is infinitely more important, but this information is rarely reported in the training literature regarding the web. This is an area virtually untouched by published research and yet vitally important if training online is to be used seriously by organizations.

RQP: Differences in Overall Effectiveness/Reactions and Moderators

The primary research question asked by this thesis is, "Are there overall differences in effectiveness and reactions between face-to-face instruction and web-based learning, and what other factors might moderate this difference?"

Each of the six specific research questions above helps address this question. There are clearly differences, although the nature of those differences is difficult to determine. People choosing online courses over traditional courses appear to be from different populations (Table 23). Older individuals tend to be enrolled in online training programs, and age may be correlated with experience, knowledge, confidence, or any number of other psychological variables, mostly unexplored in the current research literature. These pre-existing differences may affect the effectiveness of the instructional programs in which these individuals choose to participate.

Examined either by post-training scores alone or by the gains produced within-person, traditional and web-based training programs are differentially effective. However, depending upon what type of outcome is important, the superior system differs. If ultimate outcomes are of interest, measured here by post-training scores alone, it appears that online courses as designed produce equivalent outcomes for training aimed at improving knowledge and attitudes, and poorer outcomes for observable skills and problem solving skills (Table 29). Furthermore, trainees think they learn less when they complete an online course and are also less satisfied with that course (Table 20). However, if within-person gains are more important, traditional courses may be preferred for knowledge (Table 12) and problem solving skills (Table 16), with insufficient information to produce answers for the other capabilities.

As for moderators, the year that the course was designed appears to be important, with later courses generally producing better outcomes compared to traditional courses than earlier courses (Table 27). Students versus employees as trainees seems important as well, producing some of the largest differences between moderator levels found in this study (Table 26), with employees seemingly benefiting more from traditional training and students, from online training. Lab versus field produces large differences as well, although most research is conducted in the field (Table 26). Undergraduate versus graduate students as trainees seems to make a big difference as well, potentially suggesting that high complexity capabilities are more difficult to train online than are lower complexity ones (Table 28).

Conclusions

It should be noted that although this study treats quasi-experimentation as a problem, this is not necessarily the case. Appropriate use of pre-tests can help mitigate the effects of non-random assignment to produce important, meaningful findings from quasi-experimental field research. Although most of the research currently does incorporate suboptimal designs, it is the hope of the author that future research in this domain will seek to avoid such problems.

The work of Sitzmann et al. (2006) represents a solid first attempt at summarizing this quite messy and inconsistent literature, and it is the hope of the author that this thesis not be interpreted as an attack on this work. Instead, the present study simply improves upon this original effort. Sitzmann et al. (2006) concluded from their work on post-training scores alone that web-based instruction and classroom instruction were almost equally effective when similar methods were used across media. They specifically concluded that increased learner control, practice, and feedback could be used to increase the value of web-based instruction. The use of learner control, in particular, showed great gains; when examining web-based programs with high learner control in comparison to classroom instruction where learner control is low by design, they found web-based instruction produced substantially better outcomes than classroom instruction (d = .30; p. 643).

In the present study, after pre-training differences were taken into account, the analysis of this moderator came out very differently. In Table 31, it appears that learner control has a detrimental effect, producing superior outcomes when it is not present (d_{LCYes} = -.19; d_{LCNo} = .06; Δd = -.25). This highlights the importance of considering the effects of quasi-experimentation carefully. In a literature where quasi-experimentation pre-test differences are essentially random, meta-analysis would effectively nullify any artificial differences introduced by non-random assignment. In other words, pre-test differences should average to 0 across many studies, and mean post-tests alone would still be informative. In this literature however, self-selection into condition reflects a real underlying difference in populations. Probing into this underlying variable in the present study, along with roughly a 32% increase in the number of studies examined (from 96 to 126), adds substantially and substantively to the general understanding of the web-based training literature beyond the original work of Sitzmann et al. (2006).

But as with many meta-analyses, this study raises as many questions as it answers. For every difference explored, too few studies were found to fill in several cells worthy of analysis, speaking to the need for more solid, empirical research in this area. Training on the web is the likely future of much work-related training, regardless of its effectiveness. Examinations like this can be used to maximize that effectiveness, but more quality primary research is needed to fill in the gaps.

An important concern not addressed here is that of the nature of the criterion. In several of these studies, the course was developed in response to a perceived need to move a pre-existing course online. Such a design effort might introduce criterion contamination; in a sense, these course designers might be "teaching to the test" by designing their course with a specific test outcome already in mind. Some studies addressed this by having a course designer create both simultaneously, but many did not. It is ultimately unclear to what degree this kind of design effort is present in this literature, echoing Clark's (1983) early concerns.

Furthermore, testing itself varies in these courses. Some examined learning outcomes in person, while others examined learning outcomes online, and surprisingly, this was not always consistent with the course type. For example, one study might have required both its in-person and online trainees to complete learning outcome measures online, while another might split them such that online learners test online and traditional learners test in-person. Thus, the frequency of cheating may have also varied across these studies; often, too little information was presented to be clear on where testing actually took place.

From an applied perspective, several findings here are actionable for the training practitioner. The use of learner control in fully online courses does not appear to be wise; online training utilizing learner control tends to produce poorer outcomes. Hybrid courses also appear to produce superior results to traditional courses; thus, if an online training infrastructure already exists in an organization, it appears wise in terms of learning outcomes for training designers to always put supplemental material online. If that training infrastructure does not already exist, then the costs of implementing such a system need to be considered.

Regarding specific criteria, it appears that online instruction is more effective than traditional instruction when seeking knowledge and problem solving gains, equally effective when seeking attitude changes, and less effective when seeking observable skill gains. But an important issue to discuss here is that this meta-analysis in no way suggests that online instruction is *universally* preferable for these outcomes. It is tempting to make the following interpretation: "if knowledge is to be trained, online instruction should be used;" but this is risky. Instead, it should be more accurately interpreted as "studies on the training of knowledge tend to show larger gains for online in comparison to traditional instruction." Thus, special care should be made to consider the concept of Maslow's hammer: when all you have is a hammer, everything looks like a nail. Online instruction should not be considered universally superior for all knowledge outcomes; instead, in a real organization considering online training adoption, a complete needs assessment should be conducted, and the suitability of online instruction for the specific training needs of the organization should be considered carefully before committing to any particular instructional medium, online or otherwise.

At the very least, it can be concluded with a great deal of confidence that web-based instruction is generally as effective as traditional instruction. Between-subject effect sizes were universally small in comparison to the within-subject effect sizes in those same studies. Of all the findings in this thesis, the strongest is that web-based instruction *can* be roughly as effective as traditional instruction regardless of trainee, criterion, or any other moderator examined. The web is definitively a solid platform for training; the question that remains is how to make the most of it.

References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*, 3-27.

Allen, I. E., & Seaman, J. (2006). Making the grade: Online education in the United States, 2006. Retrieved October 22, 2008, from http://www.sloan-c.org/publications/survey/pdf/making_the_grade.pdf

Alzafiri, F. M. (2001). *An experimental investigation on the effects of Web-based instruction/training on cognitive and psychomotor learning.* Unpublished doctoral dissertation, University of North Texas.

American Society for Training and Development. (2008). State of the industry report: ASTD's annual review of trends in workplace learning and performance. Retrieved January 11, 2009, from http://www.astd.org/content/research/stateOfIndustry.htm

Arbaugh, J. (2005). Is there an optimal design for on-line MBA courses? *Academy of Management Learning and Education, 42*, 135-149.

Arvan, L., Ory, J. C., Bullock, C. D., Burnaska, K. K., & Hanson, M. (1998). The SCALE Efficiency Projects. *Journal of Asynchronous Learning Networks, 2*, 33-60.

Barker, B. S. (2002). *A comparison research model for adult distributed training.* Unpublished doctoral dissertation, University of Nebraska, Lincoln.

Campbell, J. P., & Kuncel, N. R. (2001). Individual and team training. In N. Anderson, D. S. Ones, H. K. Sinangil & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (Vol. 1, pp. 278-312). London, UK: Sage.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*, 445-459.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*, 21-29.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science, 210*, 207-210.

Cook, T. D., Campbell, D. T., & Peracchio, L. (1990). Quasi experimentation. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (Vol. 1). Palo Alto, CA: Consulting Psychologists Press.

Dunnette, M. D. (1966). Fads, fashions, and folderol in psychology. *American Psychologist, 21*, 343-352.

Fisher, K. M., & MacWhinney, B. (1976). AV autotutorial instruction: A review of evaluative research. *AV Communication Review, 24*, 229-261.

Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations*. Belmont, CA: Wadsworth.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledgge. *International Journal of Selection and Assessment, 8*, 275-292.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.

Jonassen, D. H., Campbell, J. P., & Davidson, M. E. (1994). Learning with media: Restructuring the debate. *Educational Technology Research and Development, 42*, 31-39.

Kopf, D. (2006). e-Learning Market to hit $52.6B by 2010. *T.H.E. Journal.* Retrieved October 22, 2008, from http://www.thejournal.com/articles/21046

Kozma, R. B. (1991). Learning with media. *Review of Educational Research, 61*, 179-211.

Kozma, R. B. (1994a). A reply: Media and methods. *Educational Technology Research and Development, 42*, 11-14.

Kozma, R. B. (1994b). Will media influence learning? Reframing the debate. *Educational Technology Research and Development, 42*, 7-19.

Kraiger, K., Ford, J., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*, 311-328.

Kulik, C.-L. C., Kulik, J. A., & Cohen, P. A. (1980). Instructional technology and college teaching. *Teaching of Psychology, 7*, 199-205.

Kulik, J. A., Bangert, R. L., & Williams, G. W. (1983). Effects of computer-based teaching on secondary school students. *Journal of Educational Psychology, 75*, 19-26.

Kulik, J. A., Kulik, C.-L. C., & Cohen, P. A. (1979). Research on audio-tutorial instruction: A meta-analysis of comparative studies. *Research in Higher Education, 11*, 321-341.

Morrison, G. R. (1994). The media effects question: "Unresolvable" or asking the right question. *Educational Technology Research and Development, 42*, 41-44.

Nielsen Company. (2009). Work panel: Weekly web usage data. Retrieved January 14, 2009, from http://www.nielsen-online.com/reports.jsp?section=pub_reports&report=usage&period=weekly&panel_type=3

Postlethwait, S. N., Novak, J., & Murry, H. T. (1971). *The audio-tutorial approach to learning*. Minneapolis: Burgess.

Reiser, R. A. (1994). Clark's invitation to the dance: An instructional Designer's response. *Educational Technology Research and Development, 42*, 45-48.

Schmeeckle, J. M. (2000). Online training: An evaluation of the effectiveness and efficiency of training law enforcement personnel over the internet (Doctoral dissertation, University of Nebraska, 2000). *Dissertation Abstracts International, 61*, 1122.

Shrock, S. A. (1994). The media influence debate: Read the fine print, but don't lose sight of the big picture. *Educational Technology Research and Development, 42*, 49-53.

Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2006). The comparative effectiveness of web-based and classroom instruction: A meta-analysis. *Personnel Psychology, 53*, 623-664.

Stadtlander, L. M. (1998). Virtual instruction: Teaching an online graduate seminar. *Teaching of Psychology, 25*, 146-148.

U.S. Census Bureau. (2005). Computer and internet use in the United States: 2003. Retrieved January 14, 2009, from http://www.census.gov/prod/2005pubs/p23-208.pdf

Wortman, P. M. (1992). Lessons from the meta-analysis of quasi-experiments. In F. B. Bryant, J. Edwards, R. S. Tindale, E. J. Posavac, L. Heath, E. Henderson & Y. Suarez-Balcazar (Eds.), *Methodological issues in applied social psychology* (pp. 65-81). New York: Plenum Press.

Meta-Analytic References

Aberson, C. L., Berger, D. E., Healy, M. R., Kyle, D. J., & Romero, V. L. (2000). Evaluation of an interactive tutorial for teaching the Central Limit Theorem. *Teaching of Psychology, 27*, 289-291.

Aberson, C. L., Berger, D. E., Healy, M. R., & Romero, V. L. (2003). Evaluation of an interactive tutorial for teaching hypothesis testing concepts. *Teaching of Psychology, 30*, 75-78.

Alonso, F., Manrique, D., & Viñes, J. M. (*in press*). A moderate constructivist e-learning instructional model evaluated on computer specialists. *Computers & Education*.

Alzafiri, F. M. (2001). *An experimental investigation on the effects of Web-based instruction/training on cognitive and psychomotor learning.* Unpublished doctoral dissertation, University of North Texas.

Arvan, L., Ory, J. C., Bullock, C. D., Burnaska, K. K., & Hanson, M. (1998). The SCALE efficiency projects. *Journal of Asynchronous Learning Networks, 2*, 33-60.

Barker, B. S. (2002). *A comparison research model for adult distributed training.* Unpublished doctoral dissertation, University of Nebraska.

Bartini, M. (2008). An empirical comparison of traditional and web-enhanced classrooms. *Journal of Instructional Psychology, 35*, 3-11.

Beile, P. M., & Boote, D. N. (2002). Library instruction and graduate professional development: Exploring the effect of learning environments on self-efficacy and learning outcomes. *Alberta Journal of Educational Research, 48*, 364-367.

Benedict, J. O., & Anderton, J. B. (2004). Applying the just-in-time teaching approach to teaching statistics. *Teaching of Psychology, 31*, 197-199.

Bliwise, N. G. (2005). Web-based tutorials for teaching introductory statistics. *Journal of Educational Computing Research 33*, 309-325.

Browning, J. B. (1999). *Analysis of concepts and skills acquisition differences between Web-delivered and classroom-delivered undergraduate instructional technology courses.* Unpublished doctoral dissertation, North Carolina State University.

Bruning, R., Siwatu, K. O., Liu, X., PytlikZillig, L. M., Horn, C., Sic, S., et al. (2008). Introducing teaching cases with face-to-face and computer-mediated discussion: Two multi-classroom quasi-experiments. *Contemporary Educational Psychology, 33*, 299-326.

Buerk, J. P., Malmstrom, T., & Peppers, E. (2003). Learning environments and learning styles: Non-traditional student enrollment and success in an Internet-based versus a lecture-based computer science course. *Learning Environments Research, 6*, 137-155.

Buzhardt, J., & Semb, G. (2005). Integrating online instruction in a college classroom to improve cost effectiveness. *Teaching of Psychology, 32*, 63-66.

Cahill, D., & Catanzaro, D. (1997). Teaching first-year Spanish on-line. *Calico Journal, 14*, 97-114.

Campbell, M. C., Floyd, J., & Sheridan, J. B. (2002). Assessment of student performance and attitudes for courses taught online versus onsite. *Journal of Applied Business Research, 18*, 45-52.

Cardin, L. H. L. (2002). *Effects of individualized web-based instructional modules on the strategic learning strategies of students enrolled in a university study skills course.* Unpublished doctoral dissertation, Texas A&M University.

Carey, J. M. (2001). Effective student outcomes: A comparison of online and face-to-face delivery modes. Retrieved November 13, 2008, from http://teleeducation.nb.ca/content/pdf/english/DEOSNEWS 11.9 effective-studentoutcomes.pdf

Casanova, R. S. (2001). *Student performance in an online general college chemistry course.* Unpublished doctoral dissertation, University of North Carolina at Wilmington.

Caulfield, J. L. (2002). *Examining the effect of teaching method and learning style on work performance for practicing home care clinicians.* Unpublished doctoral dissertation, Marquette University.

Cavus, N. (2007). Assessing the success rate of students using a learning management system together with a collaborative tool in web-based teaching of programming languages. *Journal of Educational Computing Research, 36*, 301-321.

Chan, C.-C. (2008). *Learning styles, multimedia hybrid versus traditional teaching, course satisfaction, and learning outcomes in art appreciation courses.* Unpublished doctoral dissertation, Lynn University.

Chen, C. C., Ryan, T., & Olfrnan, L. (2006). Online behavior modeling: An effective and affordable software training method. *International Journal of Web-Based Learning and Teaching Technologies, 1*, 36-53.

Chen, C. C., & Shaw, R. (2006). Online synchronous vs. asynchronous software training through the behavioral modeling approach: A longitudinal field experiment. *International Journal of Distance Education Technologies, 4*, 88-102.

Cheng, H. C. (1991). Comparison of performance and attitude in traditional and computer conferencing classes. *American Journal of Distance Education, 5*, 51-64.

Collins, M. (2000). Comparing Web, correspondence and lecture versions of a second-year non-major biology course. *British Journal of Educational Technology, 31*, 21-27.

Connolly, T. M., MacArthur, E., Stansfield, M., & McLellan, E. (2007). A quasi-experimental study of three online learning courses in computing. *Computers & Education, 49*, 345-359.

Davies, R. S., & Mendenhall, R. (1998). *Evaluation comparison of online and classroom instruction for HEPE 129--Fitness and Lifestyle Management course* (Unpublished manuscript). Provo, UT: Brigham Young University.

Day, J. A. (2008). *Investigating learning with web lectures.* Unpublished doctoral dissertation, Georgia Institute of Technology.

Dellana, S. A., Collins, W. H., & West, D. (2000). On-line education in a management science course: Effectiveness and performance factors. *Journal of Education for Business, 76*, 43-47.

Demirci, N. (2007). University students' perceptions of web-based vs. paper-based homework in a general physics course. *Eurasia Journal of Mathematics, Science & Technology Education, 3*, 29-34.

Derouza, E., & Fleming, M. (2003). A Comparison of In-Class Quizzes vs. Online Quizzes on Student Exam Performance. *Journal of Computing in Higher Education, 14*, 121-134.

DiBartola, L. M., Miller, M. K., & Turley, C. L. (2001). Do learning style and learning environment affect learning outcome? *Journal of Allied Health, 30*, 112-115.

Dinov, I. D., Sanchez, J., & Christou, N. (2008). Pedagogical utilization and assessment of the statistic online computational resource in introductory probability and statistics courses. *Computers & Education, 50*, 284-300.

Donavant, B. W. (2008). *Efficacy of distance learning for professional development of police officers.* Unpublished doctoral dissertation, University of Southern Mississippi.

Dufresne, R., Mestre, J., Hart, D. M., & Rath, K. A. (2002). The effect of web-based homework on test performance in large enrollment introductory physics courses. *Journal of Computers in Mathematics and Science Teaching, 21*, 229-251.

Ebner, M., & Holzinger, A. (2007). Successful implementation of user-centered game based learning in higher education: An example from civil engineering. *Computers & Education, 49*, 873-890.

Eppler, M. A., & Ironsmith, M. (2004). PSI and distance learning in a developmental psychology course. *Teaching of Psychology, 31*, 131-134.

Fallah, M. H., & Ubell, R. (2000). Blind scores in a graduate test: Conventional compared with web-based outcomes. *ALN Magazine, 4*, 1-5.

Fleetwood, J., Vaught, W., Feldman, D., Gracely, E., Kassutto, Z., & Novack, D. (2000). MedEthEx Online: A computer-based learning program in medical ethics and communication skills. *Teaching and Learning in Medicine, 12*, 96-104.

Fortune, M. F., Shifflett, B., & Sibley, R. E. (2006). A Comparison of online (high tech) and traditional (high touch) learning in business communication courses in Silicon Valley. *The Journal of Education for Business, 81*, 210-214.

Francescato, D., Mebane, M., Porcelli, R., Attanasio, C., & Pulino, M. (2007). Developing professional skills and social capital through computer supported collaborative learning in university contexts. *International Journal of Human-Computer Studies, 65*, 140-152.

Francescato, D., Porcelli, R., Mebane, M., Cuddetta, M., Klobas, J., & Renzi, P. (2006). Evaluation of the efficacy of collaborative learning in face-to-face and computer-supported university contexts. *Computers in Human Behavior, 22*, 163-176.

Fredda, J. V. (2000). Comparison of Selected Student Outcomes for Internet-and Campus-Based Instruction at the Wayne Huizenga Graduate School of Business and Entrepreneurship: ERIC Document Reproduction Service No. ED453743.

Friday, E., Friday-Stroud, S. S., Green, A. L., & Hill, A. Y. (2006). A multi-semester comparison of student performance between multiple traditional and online sections of two management courses. *Journal of Behavioral and Applied Management, 8*, 66-81.

Frisbie, S. H. (2002). *Sexual harassment: A comparison of online versus traditional training methods.* Unpublished doctoral dissertation, Texas Technical University, Lubbock, TX.

Frobinger, M. Q. (2008). A properly-controlled quasi-experimental examination of traditional versus online, hybrid, and web-facilitated courses that reports the reliability of all measured study variables. *Theoretical Journal of Well-Designed E-Learning Studies, 1,* 1-67.

Glenn, A. S. (2001). A comparison of distance learning and traditional learning environments: ERIC Document Reproduction Service No. ED457778.

Grabe, M., Christopherson, K., & Douglas, J. (2005). Providing introductory psychology students access to online lecture notes: The relationship of note use to performance and class attendance. *Journal of Educational Technology Systems, 33*, 295-308.

Guan, Y. S. (2008). *Factors affecting student success in an online and face-to-face educational research class.* Unpublished doctoral dissertation, Indiana State Univeristy.

Hansen, D. E. (2008). Knowledge transfer in online learning environments. *Journal of Marketing Education, 30*, 93-105.

Hiltz, S. R. (1993). Correlates of learning in a virtual classroom. *International Journal of Man-Machine Studies, 39*, 71-98.

Hisle-Gorman, E., & Zuravin, S. (2006). Teaching social work research: A comparison of Web-based and in-class lecture methods. *Journal of Technology in Human Services, 24*, 77-93.

Hurlburt, R. T. (2001). " Lectlets" deliver content at a distance: Introductory statistics as a case study. *Teaching of Psychology, 28*, 15-20.

Johnson, M. (2002). Introductory biology "online": Assessing outcomes of two student populations. *Journal of College Science Teaching, 31*, 312-317.

Johnson, S. D., Aragon, S. R., Shaik, N., & Palma-Rivas, N. (2000). Comparative analysis of learner satisfaction and learning outcomes in online and face-to-face learning environments. *Journal of Interactive Learning Research, 11*, 29-49.

Kekkonen-Moneta, S., & Moneta, G. B. (2002). E-Learning in Hong Kong: Comparing learning outcomes in online multimedia and lecture versions of an introductory computing course. *British Journal of Educational Technology, 33*, 423-433.

Keulers, B. J., Welters, C. F., Spauwen, P. H., & Houpt, P. (2007). Can face-to-face patient education be replaced by computer-based patient education? A randomised trial. *Patient Education and Counseling, 67*, 176-182.

King, F. B. (2001). *Asynchronous distance education courses employing web-based instruction: Implications of individual study skills self-efficacy and self-regulated learning.* Unpublished doctoral dissertation, University of Connecticut.

Klass, G., & Crothers, L. (2000). An experimental evaluation of Web-based tutorial quizzes. *Social Science Computer Review, 18*, 508-515.

Koory, M. A. (2003). Differences in learning outcomes for the online and F2F versions of "An Introduction to Shakespeare." *Journal of Asynchronous Learning Networks, 7*, 18-35.

Leasure, A. R., Davis, L., & Thievon, S. L. (2000). Comparison of student outcomes and preferences in a traditional vs. World Wide Web-based baccalaureate nursing research course. *The Journal of Nursing Education, 39*, 149.

Liao, C., & Chuang, S.-H. (2007). Assessing the effect of cognitive styles with different learning modes on learning outcome. *Perceptual and Motor Skills, 105*, 184-190.

Lim, J., Kim, M., Chen, S. S., & Ryder, C. E. (2008). An empirical investigation of student achievement and satisfaction in different learning environments. *Journal of Instructional Psychology, 35*, 119-119.

Liou, H. C. (1997). The Impact of WWW Texts on EFL Learning. *Computer Assisted Language Learning, 10*, 455-478.

Logan, E., & Conerly, K. (2002). Students creating community: An investigation of student interactions in a Web-based distance learning environment. Retrieved November 15, 2008

from http://www.icte.org/T01 Library/T01 253.pdf

Long, L., & Javidi, A. (2001). A comparison of course outcomes: Online distance learning versus traditional classroom settings. Retrieved November 15, 2008, from http://www.communication.ilstu.edu/activities/NCA2001/paper distance learning.pdf

Lowry, A. E. (2007). *Effects of online versus face-to-face professional development with a team-based learning community approach on teachers' application of a new instructional practice.* Unpublished doctoral dissertation, The Johns Hopkins University.

Macgregor, C. J. (2000). *Does personality matter: A comparison of student experiences in traditional and online classrooms.* Unpublished doctoral dissertation, University of Missouri, Columbia.

Maki, R. H., Maki, W. S., Patterson, M., & Whittaker, P. (2000). Evaluation of a Web-based introductory psychology course: I. Learning and satisfaction in on-line versus lecture courses. *Behavior Research Methods, Instruments & Computers, 32*, 230-239.

Mallory, C. (2007). *Leading distance learning: Possible differences in subjects enrolled in online and traditional instruction.* Unpublished doctoral dissertation, University of the Incarnate Word.

Mehlenbacher, B., Miller, C. R., Covington, D., & Larsen, J. S. (2000). Active and interactive learning online: A comparison of Web-basedand conventional writing classes. *IEEE Transactions on Professional Communication, 43*, 166-184.

Molidor, C. E. (2000). The development of successful distance education in social work: A comparison of student satisfaction between traditional and distance education classes. Retrieved November 15, 2008, from www.nssa.us/nssajrnl/18-1/pdf/14.pdf.

Mooneyhan, A. B. (2008). *The effects of using the Internet as a supplemental delivery method on university concepts of fitness classes.* Unpublished doctoral dissertation, University of Arkansas.

Musgrove, A. T. (2002). *An examination of the Kolb LSI and GEFT and their relationship to academic achievement in web-based and face-to-face nursing courses.* Unpublished doctoral dissertation, Florida Atlantic University.

Mylona, Z.-H. (1999). *Factors affecting enrollment satisfaction and persistence in a Web-based video-based and conventional instruction.* Unpublished doctoral dissertation, University of Southern California.

Navarro, P., & Shoemaker, J. (2000). Performance and perceptions of distance learners in cyberspace. *American Journal of Distance Education, 14*, 15-36.

Pankey, R. B. (1998). Piloting exercise physiology in the Web-based environment. *THE Journal (Technological Horizons In Education), 26*, 62-63.

Parker, D., & Gemino, A. (2001). Inside online learning: Comparing conceptual and technique learning performance in place-based and ALN formats. *Journal of Asynchronous Learning Networks, 5*, 64-74.

Partrich, D. E. (2004). *An analysis of learning style and grade achievement in relation to web-based and on-campus courses.* Unpublished doctoral dissertation, Southwestern Baptist Theological Seminary.

Petracchi, H., Mallinger, G., Engel, R., Rishel, C. W., & Washburn, C. (2005). Evaluating the efficacy of traditional and web-assisted instruction in an undergraduate social work practice class. *Journal of Technology in Human Services, 23*, 299-310.

Piccoli, G., Ahmad, R., & Ives, B. (2001). Web-based virtual learning environments: A research framework and a preliminary assessment of effectiveness in basic IT skills training. *MIS Quarterly, 25*, 401-426.

Poirier, C. R., & Feldman, R. S. (2004). Teaching in cyberspace: Online versus traditional instruction using a waiting-list experimental design. *Teaching of Psychology, 31*, 59-62.

Pucel, D. J., & Stertz, T. F. (2005). Effectiveness of and student satisfaction with web-based compared to traditional in-service teacher education courses. *Journal of Industrial Teacher Education, 42*, 7-23.

Ragan, R. E., & Kleoppel, J. W. (2004). Comparison of outcomes on like exams administered to in-residence and asynchronous distance-based Pharm. D. students. *Journal of Asynchronous Learning Networks, 8*, 15-24.

Reisetter, M., Lapointe, L., & Korcuska, J. (2007). The impact of altered realties: Implications of online delivery for learners' interactions, expectations, and learning skills. *International Journal on E-Learning 6*, 55-80.

Rivera, J. C., & Rice, M. L. (2002). A comparison of student outcomes and satisfaction between traditional and web based course offerings [Electronic Version]. *Online Journal of Distance Learning Administration*, 5. Retrieved November 15, 2008 from http://www.westga.edu/~distance/ojdla/fall53/rivera53.html.

Ross, J. L. (2000). *An exploratory analysis of post-secondary student achievement comparing a Web-based and a conventional course learning environment.* Unpublished doctoral dissertation, University of Calgary.

Ruiz, J. G., Smith, M., Rodriguez, O., van Zuilen, M. H., & Mintzer, M. J. (2007). An interactive e-learning tutorial for medical students on how to conduct the performance-oriented mobility assessment. *Gerontology & Geriatrics Education 28*, 51-60.

Sain, R., & Brigham, T. A. (2003). The effect of a threaded discussion component on student satisfaction and performance. *Journal of Educational Computing Research, 29*, 419-430.

Salyers, V. L. (2005). Web-enhanced and face-to-face classroom instructional methods: Effects on course outcomes and student satisfaction. *International Journal of Nursing Education Scholarship (IJNES), 2*, 0-11.

Salyers, V. L. (2007). Teaching psychomotor skills to beginning nursing students using a Web-enhanced approach: A quasi-experimental study. *International Journal of Nursing Education Scholarship (IJNES), 4*, No Pagination Specified.

Sankaran, S. R., Sankaran, D., & Bui, T. X. (2000). Effect of student attitude to course format on learning performance: An empirical study in Web vs. lecture instruction. *Journal of Instructional Psychology, 27*, 66-73.

Scheetz, N. A., & Gunter, P. L. (2004). Online versus traditional classroom delivery of a course in manual communication. *Exceptional Children, 71*, 109-120.

Schmeeckle, J. M. (2000). *Online training: An evaluation of the effectiveness and efficiency of training law enforcement personnel over the internet.* Unpublished doctoral type, University of Nebraska, Lincoln.

Schmidt, K. (2002). Classroom action research: A case study assessing students' perceptions and learning outcomes of classroom teaching versus on-line teaching. *Journal of Industrial Teacher Education, 40*, 45-59.

Schoenfeld-Tacher, R., McConnell, S., & Graham, M. (2001). Do no harm: A comparison of the effects of on-line vs. traditional delivery media on a science course. *Journal of Science Education and Technology, 10*, 257-265.

Schulman, A. H., & Sims, R. L. (1999). Learning in an online format versus an in-class format: An experimental study. *THE Journal (Technological Horizons In Education), 26*.

Sexton, J. S., Raven, M. R., & Newman, M. E. (2002). A comparison of traditional and world wide web methodologies, computer anxiety and higher order thinking skills in the in-service training of Mississippi 4-H Extension agents. *Journal of Agricultural Education, 43*, 25-36.

Shafer, M. E., Lahner, J. M., Calderone, W. K., Davis, J. E., & Petrie, T. A. (2002). The use and effectiveness of a web-based instructional program in a college student success program. *Journal of College Student Development 43*, 751-757.

Shanker, M., & Hu, M. Y. (2006). A framework for distance education effectiveness: An illustration using a business statistics course. *International Journal of Web-Based Learning and Teaching Technologies, 1*, 1-17.

Sholomskas, D. E., Syracuse-Siewert, G., Rounsaville, B. J., Ball, S. A., Nuro, K. F., & Carroll, K. M. (2005). We don't train in vain: A dissemination trial of three strategies of training clinicians in cognitive-behavioral therapy. *Journal of Consulting and Clinical Psychology, 73*, 106-115.

Sinclair, J. A. (2006). *Establishing a new paradigm in engineering and technology education: An experimental analysis of multiple learning methodologies and examination of cognitive profiles of continuing education students.* Unpublished doctoral dissertation, Nova Southeastern University.

Smith, S. B. (2000). *The effectiveness of traditional instructional methods in an online learning environment.* Unpublished doctoral dissertation, University of Nevada.

Solimeno, A., Mebane, M. E., Tomai, M., & Francescato, D. (2008). The influence of students and teachers characteristics on the efficacy of face-to-face and computer supported collaborative learning. *Computers & Education, 51*, 109-128.

Spires, H. A., Mason, C., Crissman, C., & Jackson, A. (2001). Exploring the academic self within an electronic mail environment. *Research and Teaching in Developmental Education, 17*, 5-14.

Spiros, R. K. (2003). *Individual differences in motivation during distance training: The influence of goal orientation and self-efficacy on learning outcomes.* Unpublished doctoral dissertation, The George Washington University.

Stadtlander, L. M. (1998). Virtual instruction: Teaching an online graduate seminar. *Teaching of Psychology, 25*, 146-148.

Steimberg, Y., Ram, J., Nachmia, R., & Eshel, A. (2006). An online discussion for supporting students in preparation for a test. *Journal of Asynchronous Learning Networks, 10*, 117-128.

Stephenson, J. E., Brown, C., & Griffin, D. K. (2008). Electronic delivery of lectures in the university environment: An empirical comparison of three delivery styles. *Computers & Education 50*, 640-651.

Tang, M., & Byrne, R. (2007). Regular versus online versus blended: A qualitative description of the advantages of the electronic modes and a quantitative evaluation. *International Journal on E-Learning, 6*, 257-266.

Teal, B. D. (2008). *A comparative analysis of modes of instruction using student test scores in developmental mathematics.* Unpublished doctoral dissertation, Morgan State University.

Theodorou, E. (2007). *Comparing the effects of learning the structure strategy via web-based training or classroom training on the recall of near and far transfer texts.* Unpublished doctoral dissertation, The Pennsylvania State University.

Tucker, S. (2001). Distance education: Better, worse, or as good as traditional education? [Electronic Version]. *Online Journal of Distance Learning Administration*. Retrieved November 15, 2008 from www.westga.edu/¡distance/ojdla/ winter44/tucker44.html.

Turner, M. K., Simon, S. R., Facemyer, K. C., Newhall, L. M., & Veach, T. L. (2006). Web-based learning versus standardized patients for teaching clinical diagnosis: A randomized, controlled, crossover trial. *Teaching and Learning in Medicine, 18*.

Vessell, D. C. (2001). *Comparing the WWW and WebCT to traditional methods of supporting an undergraduate psychology course: Is it effective?* Unpublished doctoral dissertation, University of Missouri, Columbia.

Wang, A. Y., & Newlin, M. H. (2000). Characteristics of students who enroll and succeed in psychology web-based classes. *Journal of Educational Psychology, 92*, 137-143.

Waschull, S. B. (2001). The online delivery of psychology courses: Attrition, performance, and evaluation. *Teaching of Psychology, 28*, 143-146.

Weems, G. H. (2002). Comparison of beginning algebra taught onsite versus online. *Journal of Developmental Education, 26*, 10-18.

Wegner, S. B., Holloway, K. C., & Garton, E. M. (1999). The effects of Internet-based instruction on student learning. *Journal of Asynchronous Learning Networks, 3*, 98-106.

Wehrs, V. H.-v., Pfafflin, M., & May, T. W. (2007). E-learning courses in epilepsy: Concept, evaluation, and experience with the E-learning course "Genetics of Epilepsies." *Epilepsia, 48*, 872-879.

White, S. E. (1999). *The effectiveness of web-based instruction: A case study.* Paper presented at the Central States Communication Association Conference, St. Louis, Missouri.

Wideman, H., & Owston, R. D. (1999). Internet-based courses at Atkinson College: An initial assessment [Electronic Version]. *Center for the Study of Computers in Education*. Retrieved November 15, 2008 from http://www.yorku.ca/irlt/ reports/techreport99-1.htm.

Wilhite, J. M. (2004). Internet versus live: Assessment of government documents bibliographic instruction. *Journal of Government Information, 30*, 561-574.

Wilson, S. P., & Harris, A. (2002). Evaluation of The Psychology Place: A web-based instructional tool for psychology courses. *Teaching of Psychology, 29*, 165-168.

Woo, M. A., & Kimmick, J. V. (2000). Comparison of Internet versus lecture instructional methods for teaching nursing research. *Journal of Professional Nursing, 16*, 132-139.

Woodward, D. K. (1998). Virtual-recitation: A World Wide Web based approach to active learning in clinical pharmacokinetics. *American Journal of Pharmaceutical Education, 62*, 329-332.

Table 1

Interrater Agreement for Study Variables

Study Variable	% Agreement	Cohen's κ
Year Course Designed	81.8%	N/A
Published	100.0%	100.0%
Trainee Type	100.0%	100.0%
Student Type	93.9%	90.9%
Hybridization Scope	90.9%	86.4%
Field/Lab	100.0%	100.0%
Experimental Design (E vs. QE)	100.0%	100.0%
Self-Selection	81.8%	72.7%
E-Location	69.7%	54.6%
Criterion Type	81.8%	77.3%
Age	96.9%	N/A
Gender	78.1%	N/A
Means	80.6%	N/A
Standard Deviations	86.2%	N/A
Learner Control	68.8%	53.2%
Human Interaction	65.0%	47.5%
Average	86.0%	79.6%

Note. Agreement and κ based on two raters across 26 (20%) articles.

Table 2

Outcome List by Setting

Setting	Capability Trained	Cases
Organizational	Programming	2
(Employees)	Cognitive Behavioral Therapy	1
	Disease Management	1
	Equine Training	1
	Genetics	1
	Jail Management	1
	Police Ethics	1
	Sexual Harassment	1
	Teaching Skills	1
	Time Management and Work Planning	1
Educational	Psychology	17
(Students)	Computer Usage	13
	Statistics	13
	Fitness	5
	Programming	5
	Biology	3
	Clinical Diagnosis	3
	Management	3

Nursing	3
Reading and Writing	3
Accounting	2
Adolescent Development	2
Business Communications	2
Communication Skills	2
Engineering	2
Ethics	2
Mathematics	2
Pharmacology	2
Physics	2
Social Work	2
American Government	1
Art Appreciation	1
Business Systems	1
Chemistry	1
Customer Value	1
Ecology	1
Economics	1
Education	1
Health Care	1

Table 3

Intercorrelation Matrix of All Study Variables

		1	2	3	4	5	6	7	8	9	10
1. Year	*r*	1.00									
	k	141									
2. Published	*r*	0.02	1.00								
	k	141	141								
3. Trainee Type	*r*	0.17	-0.12	1.00							
	k	137	137	137							
4. Student Type	*r*	-0.04	0.17	-	1.00						
	k	124	124	122	124						
5. Scope	*r*	0.25	0.02	0.44	0.12	1.00					
	k	139	139	135	124	139					

		1	2	3	4	5	6	7	8	9	10
6. Field/Lab	*r*	0.02	-0.20	-0.04	-0.04	0.21	1.00				
	n	140	140	136	123	138	140				
7. Experimental	*r*	0.11	-0.03	0.31	0.04	0.56	0.24	1.00			
	k	113	113	109	100	111	112	113			
8. Quasi w/o SS	*r*	-0.07	0.01	-0.19	0.05	-0.41	-0.14	-0.60	1.00		
	k	113	113	109	100	111	112	113	113		
9. Quasi w/SS	*r*	-0.02	0.01	-0.08	-0.09	-0.08	-0.07	-0.31	-0.58	1.00	
	k	113	113	109	100	111	112	113	113	113	
10. Hybrid Design	*r*	0.02	0.07	-0.14	0.07	0.28	0.29	0.29	-0.17	-0.09	1.00
	k	138	138	134	121	136	137	111	111	111	138
11. Knowledge	*r*	-0.19	0.09	0.01	-0.10	-0.11	-0.10	-0.22	0.14	0.07	-0.11
	k	109	109	105	96	107	108	92	92	92	108

		1	2	3	4	5	6	7	8	9	10
12. Observable Skill	*r*	0.07	0.07	-0.10	0.13	0.04	-0.04	0.00	-0.03	0.04	0.11
	n	109	109	105	96	107	108	92	92	92	108
13. Perception	*r*	0.10	-0.12	0.15	-0.12	0.09	-0.04	0.18	-0.03	-0.16	-0.04
	k	109	109	105	96	107	108	92	92	92	108
14. Prob. Solv. Skill	*r*	0.09	-0.02	-0.06	0.18	0.10	0.34	0.21	-0.20	0.01	0.16
	k	109	109	105	96	107	108	92	92	92	108
15. Attitudes	*r*	0.14	-0.19	-0.03	-0.05	-0.06	-0.01	-	-	-	-0.05
	k	109	109	105	96	107	108	92	92	92	108
16. E-Location	*r*	-0.11	-0.04	-0.21	-0.11	-0.57	-0.36	-0.56	0.33	0.19	-0.20
	k	127	127	123	114	125	127	104	104	104	126
17. TvFO Post d	*r*	0.05	0.32	-0.07	0.05	-0.07	-0.14	-0.03	-0.01	0.03	0.01
	k	88	88	84	72	86	87	69	69	69	85

		1	2	3	4	5	6	7	8	9	10
18. TvFO Pre d	*r*	0.33	0.09	-0.02	-0.20	0.11	-0.06	-0.03	-0.22	0.36	-0.15
	k	20	20	18	18	20	19	14	14	14	20
19. TvFO P-P T d	*r*	-0.32	0.24	-0.35	0.11	-0.32	-0.32	-0.28	0.36	-0.15	-0.35
	n	20	20	18	18	20	19	14	14	14	20
20. TvFO P-P FO d	*r*	-0.30	0.29	-0.07	0.10	-0.26	-0.36	-0.14	0.29	-0.23	-0.34
	k	20	20	18	18	20	19	14	14	14	20
21. TvH Post d	*r*	-0.08	-0.16	-0.16	-0.44	0.08	-	-0.20	0.20	-	-
	k	15	15	15	14	15	15	10	10	10	15
22. TvH Pre d	*r*	-0.58	0.78	-	-	0.58	-	0.86	-0.86	-	-
	k	4	4	4	4	4	4	3	3	3	4
23. TvH P-P T d	*r*	-0.20	0.10	-	-	0.20	-	-0.19	0.19	-	-
	k	4	4	4	4	4	4	3	3	3	4

		1	2	3	4	5	6	7	8	9	10
24. TvH P-P H d	*r*	0.11	0.10	-	-	-0.11	-	-0.79	0.79	-	-
	k	4	4	4	4	4	4	3	3	3	4
25. TvWF Post d	*r*	-0.06	0.42	-	0.13	0.07	-	0.31	-0.13	-0.19	0.16
	k	29	29	29	29	29	29	24	24	24	29
26. TvWF Pre d	*r*	-0.48	0.59	-	-	-	-	0.73	-0.71	0.17	0.43
	n	7	7	7	7	7	7	6	6	6	7
27. TvWF P-P T d	*r*	0.46	0.59	-	-	-	-	-0.12	-0.41	0.64	-0.26
	k	7	7	7	7	7	7	6	6	6	7
28. TvWF P-P WF d	*r*	0.45	0.60	-	-	-	-	-0.09	-0.63	0.89	-0.14
	k	7	7	7	7	7	7	6	6	6	7
29. Learn. Con. FO	*r*	-0.25	0.09	-0.07	-0.32	-0.42	-0.28	-0.06	-0.22	0.31	-0.24
	k	49	49	46	41	47	48	42	42	42	46

		1	2	3	4	5	6	7	8	9	10
30. Interaction FO	*r*	0.10	-0.07	0.14	-0.16	-0.31	-0.06	-0.17	-0.08	0.29	-0.16
	k	49	49	47	41	47	48	39	39	39	48

/

		11	12	13	14	15	16	17	18	19	20
11. Knowledge	*r*	1.00									
	k	109									
12. Observable Skill	*r*	-0.60	1.00								
	k	109	109								
13. Perception	*r*	-0.56	-0.08	1.00							
	k	109	109	109							
14. Prob. Solv. Skill	*r*	-0.39	-0.06	-0.05	1.00						
	k	109	109	109	109						
15. Attitudes	*r*	-0.19	-0.03	-0.03	-0.02	1.00					
	k	109	109	109	109	109					
16. E-Location	*r*	0.13	-0.18	0.12	-0.21	0.04	1.00				
	n	100	100	100	100	100	127				

		11	12	13	14	15	16	17	18	19	20
17. TvFO Post d	*r*	-0.02	0.21	-0.12	-0.09	-	0.07	1.00			
	k	65	65	65	65	65	77	88			
18. TvFO Pre d	*r*	0.04	0.09	-0.08	-0.04	-	-0.29	0.13	1.00		
	k	19	19	19	19	19	17	20	20		
19. TvFO P-P T d	*r*	0.17	-0.23	0.16	-0.30	-	0.35	-0.32	-0.35	1.00	
	k	19	19	19	19	19	17	20	20	20	
20. TvFO P-P FO d	*r*	0.21	-0.23	0.14	-0.35	-	0.46	0.05	-0.56	0.88	1.00
	k	19	19	19	19	19	17	20	20	20	20
21. TvH Post d	*r*	0.30	-0.01	-	-	-0.50	-0.21	1.00	-	-	-
	k	12	12	12	12	12	15	2	1	1	1
22. TvH Pre d	*r*	0.94	-	-	-	-0.94	-0.58	-	-	-	-
	n	4	4	4	4	4	4	1	1	1	1

		11	12	13	14	15	16	17	18	19	20
23. TvH P-P T d	*r*	0.86	-	-	-	-0.86	-0.20	-	-	-	-
	k	4	4	4	4	4	4	1	1	1	1
24. TvH P-P H d	*r*	0.85	-	-	-	-0.85	0.11	-	-	-	-
	k	4	4	4	4	4	4	1	1	1	1
25. TvWF Post d	*r*	0.51	-0.33	-0.23	-0.07	-0.28	-	-	-	-	-
	k	28	28	28	28	28	26	0	0	0	0
26. TvWF Pre d	*r*	0.59	-	0.20	-	-0.95	-	-	-	-	-
	k	7	7	7	7	7	5	0	0	0	0
27. TvWF P-P T d	*r*	0.59	-	-0.20	-	-0.56	-	-	-	-	-
	k	7	7	7	7	7	5	0	0	0	0
28. TvWF P-P WF d	*r*	0.60	-	-0.37	-	-0.41	-	-	-	-	-
	n	7	7	7	7	7	5	0	0	0	0

		11	12	13	14	15	16	17	18	19	20
29. Learn. Con. FO	*r*	0.20	-0.16	0.00	-0.30	-	0.43	-0.05	-0.22	0.03	0.33
	k	44	44	44	44	44	44	44	13	13	13
30. Interaction FO	*r*	-0.06	0.20	0.03	-0.18	-	0.25	-0.06	-0.13	-0.17	-0.18
	k	41	41	41	41	41	43	45	13	13	13

		21	22	23	24	25	26	27	28	29	30
21. TvH Post d	*r*	1.00									
	k	15									
22. TvH Pre d	*r*	0.73	1.00								
	k	4	4								
23. TvH P-P T d	*r*	0.58	0.70	1.00							
	k	4	4	4							
24. TvH P-P H d	*r*	0.77	0.63	0.93	1.00						
	k	4	4	4	4						
25. TvWF Post d	*r*	1.00	-	-	-	1.00					
	k	3	1	1	1	29					
26. TvWF Pre d	*r*	-	-	-	-	0.85	1.00				
	n	1	1	1	1	7	7				

		21	22	23	24	25	26	27	28	29	30
27. TvWF P-P T d	*r*	-	-	-	-	0.18	0.48	1.00			
	k	1	1	1	1	7	7	7			
28. TvWS P-P WF d	*r*	-	-	-	-	0.30	0.40	0.89	1.00		
	k	1	1	1	1	7	7	7	7		
29. Learn. Con. FO	*r*	-	-	-	-	-	-	-	-	1.00	
	k	2	0	0	0	0	0	0	0	49	
30. Interaction FO	*r*	-	-	-	-	-	-	-	-	0.22	1.00
	n	3	1	1	1	0	0	0	0	38	49

Note. Year = year of publication; Published: 1 = published, 0 = unpublished; Trainee Type: 0 = student, 1 = employee; Student Type: 0 = undergraduate, 1 = graduate; Scope: 0 = semester, 1 = specific; Field/Lab: 0 = field, 1 = lab; Experimental, Quasi w/o SS, Quasi w/SS, Knowledge, Observable Skill, Perception, Problem Solving Skill, Attitudes all dummy coded; Hybrid Design: 0 = class, 1 = addition to pre-existing class; E-Location: 0 = restricted, 1 = learner choice; TvFO = Traditional vs. Fully Online; TvH = Traditional vs.

Hybrid; TvWF = Traditiona vs. Web-facilitated; Post = comparison of post-training scores; Pre = between-person comparison of pre-training scores; P-P = within-person comparison of post- to pre-training scores; Learn. Con. = dummy coded presence of learner control in the fully online condition; Interaction = human interaction in the fully online condition

Table 4

Between-Person Differences in Instructional Outcomes

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Web-facilitated Courses								
Post-Training	5273	26	.14	.37	.35	-.31	.58	14
Pre-Training	936	6	-.03	.10	.00	-.03	-.03	100
Traditional vs. Hybrid Courses								
Post-Training	1108	11	.34	.33	.26	.00	.67	37
Pre-Training	285	3	.04	.16	.00	.04	.04	100
Traditional vs. Fully Online Courses								
Post-Training	14255	88	.17	.40	.37	-.30	.64	16
Pre-Training	1503	20	.17	.28	.15	-.02	.36	70

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 5

Between- and Within-Person Differences between Traditional and Fully Online Courses

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Between-Person								
Post-Training	14255	88	.17	.40	.37	-.30	.64	16
Pre-Training Unavailable	12751	68	.17	.39	.36	-.29	.64	14
Pre-Training Available	1504	20	.12	.46	.40	-.39	.62	25
Pre-Training	1503	20	.17	.28	.15	-.02	.36	70
Within-Person								
Traditional	820	20	2.19	1.28	1.16	1.16	3.67	19
Fully Online	686	20	1.99	1.19	1.03	.67	3.31	25

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 6

Pre-Training Between-Person Differences between Traditional and Fully Online Courses by Experimental Design

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Fully Online	1503	20	.17	.28	.15	-.02	.36	70
Experimental Design								
Experimental	328	5	.02	.04	.00	.02	.02	100
QE without Self-Selection	202	4	.01	.48	.39	-.49	.52	34
QE with Self-Selection	586	7	.19	.25	.13	.03	.35	75

Note. QE = quasi-experimental, k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 7

Between-Person Differences between Traditional and Fully Online Courses by Trainee Type

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Fully Online								
Trainee Type								
Student (Educational)								
Pre-Training	1320	16	.17	.25	.11	.03	.31	80
Post-Training	13133	73	.19	.37	.34	-.24	.62	16
Employee (Organizational)								
Pre-Training	85	2	.11	.02	.00	.11	.11	100
Post-Training	914	11	-.17	.52	.47	-.77	.43	18

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 8

Post-Training Between-Person Differences between Traditional and Fully Online Courses by Setting

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Fully Online	14255	88	.17	.40	.37	-.30	.64	16
Setting								
Lab	97	2	-.39	.03	.00	-.39	-.39	100
Field	14123	85	.17	.40	.37	-.30	.64	15

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 9

Between- and Within-Person Differences between Traditional and Hybrid Courses

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Between-Person								
Post-Training	1192	12	.31	.33	.27	-.03	.65	37
Pre-Training Unavailable	907	9	.33	.35	.29	-.04	.70	32
Pre-Training Available	285	3	.24	.25	.15	.05	.42	67
Pre-Training	285	3	.04	.16	.00	.04	.04	100
Within-Person								
Traditional	163	3	1.28	.46	.19	1.03	1.52	83
Hybrid	122	3	1.41	.34	.00	1.41	1.41	100

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 10

Between- and Within-Person Differences between Traditional and Web-facilitated Courses

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Between-Person								
Post-Training	5273	26	.14	.37	.35	-.31	.58	14
Pre-Training Unavailable	4364	20	.12	.40	.37	-.36	.59	12
Pre-Training Available	909	6	.24	.21	.13	.07	.41	61
Pre-Training	936	6	-.03	.10	.00	-.03	-.03	100
Within-Person								
Traditional	538	6	.91	.66	.58	.16	1.66	22
Web-facilitated	385	6	.85	.64	.52	.18	1.52	33

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 11

Within-Person Differences between Traditional and Fully Online Courses by Trainee Type

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Within-Person								
Student (Educational)								
Traditional	729	16	2.27	1.22	1.09	.87	3.67	19
Fully Online	607	16	2.00	1.14	.99	.74	3.26	25
Employee (Organizational)								
Traditional	41	2	.62	.30	.00	.62	.62	100
Fully Online	44	2	1.61	.87	.52	.95	2.27	64

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 12

Between- and Within-Person Differences in Knowledge Outcomes between Traditional and Fully Online Courses

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Fully Online Courses, Knowledge Criteria								
Between-Person								
Post-Training	5339	52	.27	.47	.43	-.28	.82	17
Pre-Training	1247	17	.14	.29	.17	-.07	.35	66
Within-Person								
Traditional	686	17	2.04	1.24	1.12	.61	3.47	19
Fully Online	556	17	1.88	1.11	.94	.68	3.09	29

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 13

Between- and Within-Person Differences in Knowledge Outcomes between Traditional and Hybrid Courses

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Hybrid Courses, Knowledge Criteria								
Between-Person								
Post-Training	258	3	.26	.25	.13	.09	.43	73
Pre-Training	174	2	-.07	.10	.00	-.07	-.07	100
Within-Person								
Traditional	107	2	1.23	.57	.38	.75	1.71	56
Hybrid	67	2	1.67	.27	.00	1.67	1.67	100

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 14

Between- and Within-Person Knowledge Differences between Traditional and Web-facilitated Courses

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Web-facilitated Courses, Knowledge Criteria								
Between-Person								
Post-Training	1494	10	-.05	.54	.52	-.71	.62	9
Pre-Training	338	2	-.15	.05	.00	-.15	-.15	100
Within-Person								
Traditional	193	2	.53	.01	.00	.53	.53	100
Web-facilitated	145	2	.75	.07	.00	.75	.75	100

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 15

Between-Person Differences in Observable Skill Outcomes

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Between-Person								
Traditional vs. Fully Online	1206	16	.00	.53	.47	-.60	.61	19
Traditional vs. Hybrid	144	4	.71	.37	.14	.53	.89	86
Traditional vs. Web-facilitated	457	5	.06	.36	.29	-.32	.43	34

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 16

Between- and Within-Person Differences in Problem Solving Skill Outcomes between Traditional and Fully Online Courses

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Fully Online Courses, Problem Solving Skill Criteria								
Between-Person								
Post-Training	542	8	.27	.30	.17	.04	.49	67
Pre-Training	165	2	.40	.13	.00	.40	.40	100
Within-Person								
Traditional	75	2	.83	.11	.00	.83	.83	100
Fully Online	90	2	.79	.08	.00	.79	.79	100

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 17

Between- and Within-Person Differences in Problem Solving Skill Outcomes between Traditional and Web-facilitated Courses

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Web-facilitated Courses, Problem Solving Skill Criteria								
Between-Person								
Post-Training	100	3	.68	.51	.37	.21	1.15	49
Pre-Training	64	2	.40	.18	.00	.40	.40	100
Within-Person								
Traditional	34	2	.22	.43	.00	.22	.22	100
Web-facilitated	30	2	.82	.23	.00	.82	.82	100

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 18

Between-Person Differences in Attitudes Outcomes between Traditional and Fully Online Courses

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Fully Online Courses								
Between-Person								
Post-Training	203	2	.00	.23	.11	-.15	.15	75
Pre-Training	116	1	-.21	-	-	-	-	-

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 19

Between-Person Differences in Perception Outcomes between Traditional and Fully Online Courses

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Fully Online Courses, Perception Criteria								
Between-Person								
Post-Training	789	7	-.16	.18	.00	-.16	-.16	100
Pre-Training	386	3	.09	.26	.19	-.16	.33	46
Within-Person								
Traditional	195	3	3.88	.21	.00	3.88	3.88	100
Fully Online	191	3	3.82	.45	.00	3.82	3.82	100

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 20

Between-Person Differences in Reactions to Training between Traditional and Fully Online Courses by Experimental Design

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Fully Online	2155	27	-.21	.65	.61	-.98	.57	12
Experimental Design								
Experimental	237	5	-.47	.63	.55	-1.18	.24	22
QE without Self-Selection	96	2	-.70	.01	.00	-.70	-.70	100
QE with Self-Selection	806	11	-.16	.92	.89	-1.30	.98	6

Note. QE = quasi-experimental, k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 21

Between-Person Differences in Reactions to Training between Traditional and Web-facilitated Courses by Experimental Design

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Web-facilitated	812	6	-.50	.49	.45	-1.08	.08	13
Experimental Design								
Experimental	110	1	.37	-	-	-	-	-
QE without Self-Selection	361	3	-.54	.50	.46	-1.13	.06	14
QE with Self-Selection	341	2	-.73	.01	.00	-.73	-.73	100

Note. QE = quasi-experimental, k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 22

Mapping of Pre-Training Individual Difference Measures to Category

Category	Original Scale Name
Ability	Cognitive Ability
Affect	Attitudes Towards Computers
	Computer Anxiety (-)
	Confidence
	Motivation to Succeed
	Perceived Academic Self Efficacy
	Perceived Capacity to Pursue Aims
	Perceived Resilience and Hopefulness
	Perceived Social Efficacy
	Perceived Problem Solving Efficacy
Computer/Internet Experience	Computer Access
	Computer Literacy
	Computer Usage
	E-Learning Experience
	Email Use
	Internet Experience
	Use of Computers in Class

Table 23

Differences in Pre-Training Individual Difference Categories between Traditional and Fully Online Courses from Individuals Self-Selecting into Condition

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional v. Fully Online								
Age	210	3	1.61	.08	.00	1.61	1.61	100
Gender	221	3	-.06	.16	.00	-.06	-.06	100
All Psychological Scales	342	4	.10	.11	.00	.10	.10	100
Affect	140	3	-.05	.04	.00	-.05	-.05	100
Computer/Internet Experience	255	2	.20	.08	.00	.20	.20	100

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates individual differences were lower for traditional course

Table 24

Differences in Pre-Training Individual Difference Categories between Traditional and Hybrid/Web-facilitated Courses from Individuals Self-Selecting into Condition

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional v. Web-facilitated and Hybrid								
Age	193	2	.02	.11	.00	.02	.02	100
Gender	193	2	.17	.14	.00	.17	.17	100
All Psychological Scales	465	3	-.04	.03	.00	-.04	-.04	100
Ability	160	2	-.07	.06	.00	-.07	-.07	100
Affect	465	3	-.38	.31	.27	-.72	-.04	27
Computer/Internet Experience	305	1	.48	-	-	-	-	-

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates individual differences were lower for traditional course

Table 25

Analysis of Experimental Design Moderator for Traditional vs. Fully Online Course Outcomes Using Post-Test Scores Adjusted for Quasi-Experimental Design with Self-Selection by d *= 0.19*

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Fully Online	14255	88	.08	.41	.38	-.41	.58	14
Experimental Design								
Experimental	1051	17	.03	.59	.53	-.66	.71	19
QE without Self-Selection	461	9	-.19	.37	.24	-.50	.12	57
QE with Self-Selection	6846	48	-.06	.40	.36	-.52	.41	18

Table 26

Analysis of Setting and Trainee Type Moderators for Traditional vs. Fully Online Course Outcomes Using Post-Test Scores Adjusted for Quasi-Experimental Design with Self-Selection by d *= 0.19*

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Fully Online	14255	88	.08	.41	.38	-.41	.58	14
Setting								
Lab	97	2	-.39	.03	.00	-.39	-.39	100
Field	14123	85	.08	.42	.39	-.42	.57	14
Trainee Type								
Student (Educational)	13133	73	.09	.39	.36	-.37	.56	14
Employee (Organizational)	914	11	-.20	.53	.48	-.82	.42	17

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 27

Analysis of Year Moderator for Traditional vs. Fully Online Course Outcomes Using Post-Test Scores Adjusted for Quasi-Experimental Design with Self-Selection by d *= 0.19*

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Fully Online	14255	88	.08	.41	.38	-.41	.58	14
Year of Publication								
1991-1995	213	2	-.17	.09	.00	-.17	-.17	100
1996-2000	5223	22	.06	.29	.26	-.28	.40	19
2001-2005	5323	37	.29	.40	.36	-.17	.75	18
2006-2009	3496	27	-.21	.43	.39	-.71	.29	17
Year Online Course Designed								
1996-2000	6994	26	-.01	.28	.25	-.32	.31	20
2001-2005	1520	14	.07	.62	.59	-.68	.83	10

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 28

Analysis of Student Type and Scope Moderators for Traditional vs. Fully Online Course Outcomes Using Post-Test Scores Adjusted for Quasi-Experimental Design with Self-Selection by d *= 0.19*

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Fully Online	14255	88	.08	.41	.38	-.41	.58	14
Student Type (Educational Only)								
Undergraduate	11083	55	.11	.40	.38	-.38	.59	12
Graduate	261	4	-.08	.34	.23	-.37	.21	54
Scope (Educational Only)								
Specific Intervention	996	15	-.02	.27	.10	-.15	.12	85
Semester-long Course	4152	14	.07	.27	.24	-.24	.38	19

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 29

Analysis of Criterion Type Moderators for Traditional vs. Fully Online Course Outcomes Using Post-Test Scores Adjusted for Quasi-Experimental Design with Self-Selection by d *= 0.19*

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Fully Online	14255	88	.08	.41	.38	-.41	.58	14
Criterion Type								
Knowledge	5339	52	.19	.46	.42	-.35	.73	18
Observable Skill	1206	16	-.12	.53	.48	-.74	.49	19
Problem Solving Skill	542	8	.21	.31	.19	-.04	.46	61
Attitudes	203	2	.00	.23	.11	-.15	.15	75
Perception	789	7	-.20	.24	.15	-.39	-.00	61

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 30

Analysis of Publication and E-Location Moderators for Traditional vs. Fully Online Course Outcomes Using Post-Test Scores Adjusted for Quasi-Experimental Design with Self-Selection by d *= 0.19*

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Fully Online	14255	88	.08	.41	.38	-.41	.58	14
Published								
Yes	9089	64	.12	.48	.45	-.46	.69	12
No	5166	24	.00	.26	.22	-.28	.28	28
E-Location								
Restricted	510	10	.03	.46	.36	-.44	.49	38
Learner's Choice	12513	67	.10	.42	.40	-.41	.61	12

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores

Table 31

Analysis of Course Design Moderators for Traditional vs. Fully Online Course Outcomes Using Post-Test Scores Adjusted for Quasi-Experimental Design with Self-Selection by d *= 0.19*

	N	*k*	Mean *d*	SD_{obs}	SD_{res}	CI_{lower}	CI_{higher}	% Var
Traditional vs. Fully Online	14255	88	.08	.41	.38	-.41	.58	14
Learner Control on Web								
Yes	836	15	-.19	.68	.62	-.98	.60	16
No	1055	11	.06	.33	.26	-.28	.39	37
Human Interaction on Web								
Yes	4203	27	-.10	.42	.39	-.61	.40	14
No	1290	18	-.04	.50	.44	-.61	.53	22

Note. k = number of independent effect sizes, SD_{obs} = standard deviation of observed *d*, SD_{res} = residual standard deviation, CI = 80% credibility interval lower and upper bounds, % Var = percentage variance accounted for, positive *d* indicates traditional course produced poorer (lower) scores