The Comparative Effectiveness of Web-Based and Classroom Instruction:

A Meta-Analysis

Traci M. Sitzmann
Advanced Distributed Learning
Kurt Kraiger
Colorado State University
David W. Stewart
University of Tulsa
Robert A. Wisher
Department of Defense

Running Head: Effectiveness of Web-Based Instruction
Abstract

Meta-analytic techniques were used to examine the effectiveness of Web-based instruction (WBI) relative to classroom instruction (CI) and to examine moderators of the comparative effectiveness of the two delivery media. Overall the results indicate WBI is 6% more effective than CI for teaching declarative knowledge, the two delivery media are equally effective for teaching procedural knowledge, and trainees are equally satisfied with WBI and CI. However, WBI and CI were equally effective for teaching declarative knowledge when the same instructional methods were used to deliver the two courses, suggesting media effects are spurious and supporting Clark’s (1983, 1994) theory. Finally, WBI was 14% more effective than CI for teaching declarative knowledge when trainees were provided with control during WBI and in long training courses. Study limitations and directions for future research are discussed.

Keywords:
Web-based training
Online training
Evaluation
Meta-analysis
The Comparative Effectiveness of Web-Based and Classroom Instruction:  
A Meta-Analysis

Web-based instruction (WBI) is becoming a favored training option in industry, government, and higher education. WBI is a “hypermedia-based instructional program which utilizes the attributes and resources of the World Wide Web to create a meaningful learning environment where learning is fostered and supported” (Khan, 1997, p. 6). WBI is delivered via the computer using the Internet, making it capable of instant updating, distribution, and sharing of information (Rosenberg, 2001). In a survey of organizations in the American Society of Training and Development's benchmarking service, the percentage of companies using technology-delivered training increased from 8% in 1999 to 24% in 2003, and more than half of the technology-based courses in 2003 were delivered online (Sugrue & Kim, 2004). In addition, over 1,100 institutions of higher education in the United States offer online courses (Newman & Scurry, 2001). Elsewhere, the military is pursuing a widespread adoption of WBI to meet its enormous training requirements (General Accounting Office, 2003). The Army uses online instruction as a retention tool, with over 40,000 soldiers in 50 countries pursuing advanced degrees online (Symonds, 2003). Finally, in a recent survey the majority of learning executives anticipated increasing use of online platforms to deliver higher education to their employees (Trierweller & Rivera, 2005).

Given its growing popularity, it is important to understand whether or not this delivery medium is effective, whether it is more effective than other delivery media, and what contextual or methodological factors moderate its effectiveness. In the current study, effectiveness is operationalized as both reactions to and learning from WBI. The present study examines cumulative evidence of the effectiveness of WBI relative to classroom instruction (CI) and
Effectiveness of Web-Based Instruction as an Applied Issue

The rush to implement online learning preceded empirical evidence of its benefits. Given the increasingly widespread implementation of WBI, it is important to determine whether or not WBI is effective for imparting useful knowledge and skills. As a form of technology-assisted instruction (TAI), online instruction will have utility to organizations and institutions if it results in learning and retention, is well-received by users, and is cost-effective to the sponsoring organization or institution. There have been few studies of the cost-effectiveness of online instruction, but a sufficient number of primary studies have now been conducted to determine its effectiveness with respect to learning and user reactions. However, Arbaugh (2005) has questioned whether single studies are useful for understanding the impact of technology and course characteristics on WBI effectiveness. By examining trends over studies, we can form quantified conclusions of WBI only a decade after its implementation. Given evidence that WBI is effective, more organizations and institutions will be able to justify the expenditures necessary to adopt it. If evidence suggests that it is not as effective as existing delivery media, organizations and institutions may be more cautious about replacing traditional delivery media with WBI, or develop more effective online training methods. Finally, if WBI is effective under some conditions and not others, organizations and institutions that place training online can use the results of this study to identify optimal conditions for learning.

Accordingly, this meta-analysis addresses an important applied question by examining cumulative evidence of the effectiveness of WBI relative to CI in terms of users’ learning and reactions. Thus, the current study is a meta-analysis of studies that compares the effectiveness of WBI and CI for delivering instruction on the same topic. CI was used as a basis for
comparison since it is the most common training technique (Sugrue & Kim, 2004) and because there is still a paucity of studies comparing WBI to other instructional media.

*Effectiveness of Web-Based Instruction as a Theoretical Issue*

The question of whether or not WBI is more effective than other delivery media has theoretical importance. Educational psychologist Richard Clark (1983; 1994) has been a long-time critic of studies and reviews that purport to show that newer, technologically-based instructional media are superior to existing media. While media is often used to refer to the general method of delivering training, here media refers to *technological* devices used for the purpose of instruction (Clark & Sugrue, 1995). Clark has argued that delivery media (such as computers or distance learning) are relatively inconsequential in affecting learning outcomes, compared to more powerful influences such as individual differences and instructional methods. Instructional methods refer to strategies used within a course to convey course content such as providing opportunities for practice or group discussions.

Clark (1983; 1994) criticized most media effectiveness research on two grounds. First, most studies fail to institute experimental controls sufficient to rule out alternative explanations for group differences. Second, Clark argued that most prior studies have failed to isolate instructional attributes that are *unique* to a single medium. For example, WBI may provide more opportunities for learner customization than CI, but: a) classroom learning can provide some customization in some situations; and b) opportunities for learner customization are not unique to WBI. Clark argued that if studies fail to isolate attributes unique to the medium, results of those studies cannot be accepted as evidence of the superiority of the medium. In short, Clark argued that there is nothing uniquely beneficial about any computer-aided instructional medium (including WBI).
Clark’s position has received broad support (e.g., Bernard et al., 2004; Russell, 1999), but is not without its critics. Cobb (1997) argued that certain instructional methods, while common to multiple media, are more efficient or potent in one medium than others. For example, in a film studies course, the effects of background music on viewer mood can be demonstrated much more effectively using video than via lecture or book. By extension, a flexible medium such as WBI might be more effective if it can utilize multiple instructional methods given the nature of the learning material. Kozma (1994) argued that while it may be difficult to isolate individual instructional attributes to any single medium, it is possible to identify clusters of attributes (e.g., customization and hyperlinking) that are more efficiently accomplished in one medium rather than others. For example, compared to CI, WBI is more likely to offer customization of instructional methods and content, as well as continual access.

As a special type of hypermedia learning system, WBI represents a non-linear instructional medium that may encourage deeper processing and cognitive flexibility in learners (Spiro & Jehng, 1990). Hypermedia is instruction organized in a nonlinear format that allows learners to reference material based on their individual needs and uses a wide variety of media (e.g., sound, animation, text, video) (Tessmer, 1993). The use of hypermedia may not only lead to deeper learning by encouraging trainees to think about how new information is related to existing knowledge (Salomon, 1988), but may be a superior medium to the extent that it offers a cluster of learning modes (e.g., text, audio, graphics, synchronous and asynchronous communication) that can be tailored to individual learning styles. Arbaugh (2005) also detailed clusters of features of WBI that may lead to greater instructional effectiveness including media variety, facilitation of Web exploration, and learner ease and flexibility of use.
In summary, there are two schools of thought with respect to the relative effectiveness of WBI: Clark’s position that argues no instructional medium is uniquely advantageous, and a pro-technology group that believes that WBI provides greater flexibility and greater access to multiple learning modes such that it may be superior to media that are more grounded in single instructional methods (Dumont, 1996; Hiltz & Wellman, 1997; Sullivan, 2001).

We intend to apply meta-analytic methods to address the important theoretical question of whether instructional media matters. We will do so in three ways, by examining the arguments that Clark (1983; 1994) uses to refute pro-media studies. First, we will look at a subset of all studies in which a true experimental design was used. Clark has argued that past research that has supported certain media has failed to execute proper experimental procedures that control for participant motivation or prior experience with the technology. Support for Clark’s position would come from the following pattern of results: positive mean effect sizes for learning and reactions when analyzing all studies, but no effects for media when analyzing only true experiments.

Second, we will examine a subset of studies that equate instructional methods across conditions. Clark has argued that media studies often confound media with instructional methods or content, so that it is impossible to determine whether main effects are due to differences in the medium or differences in methods or content. However, we were able to isolate studies that compared WBI to CI but used identical teaching methods (e.g., lecture and work problems). Support for the pro-technology position would be evident in greater learning or more positive reactions from WBI relative to CI even when instructional methods do not differ across groups. Support for Clark’s theory would be evident in no differences between WBI relative to CI when instructional methods are the same across courses.
Third, we will examine a subset of studies in which WBI was used alone to studies in which WBI was used as a supplement to CI (henceforth WBI-S). Although Clark has not addressed the additive effects of multiple versions of media, if media do not matter, there should be no difference in the relative effectiveness of instruction among CI alone, WBI alone, or WBI-S (provided content is identical across groups). However, support for the pro-technology position would be evident in greater learning or more positive reactions when learners are offered more choices of media as in the case of CI supplemented with WBI. Given these practical and theoretical issues, we proposed a number of research questions.

*Research Questions – Main Effects*

The first objective of the study was to examine the effectiveness of WBI relative to CI. WBI is a form of TAI in which content is available over the World-Wide Web rather than at a single work-station. Prior meta-analytic studies have reported overall positive effect sizes for various forms of TAI compared to CI including videodiscs (Fletcher, 1990), computer-assisted training (Kulik, 1994; Kulik & Kulik, 1991; Yaakub, 1998), hypermedia systems (Liao, 1999), and distance education (Machtmes & Asher, 2000; Zhao, Lei, Lai, & Tan, 2005). Accordingly, we expect similar positive effect sizes for WBI.

In addition to studying the effects of WBI alone, we were also interested in the relative effectiveness of training in which WBI was used to supplement traditional classroom approaches. In both higher education and corporate training, the use of WBI to supplement traditional face-to-face instruction is known as blended learning, and is perceived by many to a strong instructional approach that incorporates both the benefits of personal interaction and self-study between instructional meetings using the Web (Kerres & deWitt; 2003; Masie, 2002). According to media richness theory (Daft & Lengel, 1986), there are instructional benefits to
presenting training content using multiple media. However, there is not yet strong evidence of the effectiveness of this increasingly popular instructional approach.

The current meta-analysis utilizes Kraiger, Ford, & Salas’ (1993) multidimensional framework of learning: affective, cognitive, and skill-based knowledge. Cognitive learning (a.k.a. declarative knowledge) refers to information (e.g., facts or definitions) about a content domain. Skill-based knowledge (a.k.a. procedural knowledge) refers to information about how to perform a task or action. Affective learning refers to attitudes or values relevant to the objectives of the training course (e.g., appreciation for the topic area or self-efficacy). However, few studies measured affective learning so we focused exclusively on declarative and procedural knowledge. Thus, we wished to compare WBI and WBI-S to CI in terms of their effectiveness for teaching declarative and procedural knowledge. We proposed the following research questions:

Q1: Is WBI or CI more effective for teaching declarative knowledge?

Q2: Is WBI or CI more effective for teaching procedural knowledge?

Q3: Are differences in WBI learning outcomes relative to CI greater when the Web is the sole means of instruction or when the Web is used as a supplement to CI?

The second set of research questions examines the utility of WBI and WBI-S relative to CI with respect to learner reactions. It is difficult to formulate directional hypotheses about preferences for WBI, WBI-S, or CI. Learners likely gravitate towards one medium based on individual preferences. Tailoring training to the needs, prior knowledge, and interests of individual learners and the opportunity to access training material any time may result in trainees reacting more favorably towards WBI. However, CI provides face-to-face interaction with the instructor and other trainees and should circumvent frustrations that can be associated with
unreliable technology. Finally, reactions to online instruction may also vary across settings and learners depending on the perceived usefulness and ease of use of the technology (Davis, 1989). While it is difficult to predict that learners will react more positively to one medium than another, it is an interesting research question. Accordingly, we propose the following research questions:

**Q4:** *Do trainees react more favorably towards WBI or CI?*

**Q5:** *Is the difference in reactions towards WBI relative to CI greater when the Web is the sole means of instruction or when the Web is used as a supplement to CI?*

**Research Questions – Research Design Moderators**

A second objective of the study was to examine moderators of the effectiveness of WBI relative to CI. The first moderator analysis investigates the impact of WBI relative to CI when differences in instructional methods are eliminated. This investigation directly addresses the criticisms of Clark (1983; 1994), who has argued that it is instructional methods (e.g., lecture, practice, examples, discussion) that influence participant achievement while the delivery media (e.g., WBI v. CI) influence only the cost and accessibility of the information. Thus, to provide a more objective assessment of the relative effectiveness of WBI and CI, we examined differences in the effectiveness of the two delivery media when the same instructional methods were used to deliver the two courses. If Clark is correct, any observed effects for WBI should disappear when we control for instructional methods.

Second, we examined a subset of all studies that used true experimental designs. This addresses Clark’s concerns that media comparison studies often confound instructional mediums with instructional quality, subject motivation, and so forth. Random assignment of participants to WBI and CI should reduce differences between test groups that that exaggerate effects on
learning or reactions. Accordingly, any observed differences in learning and reactions can be more readily attributed to the instructional medium. If Clark’s position is correct, any observed effects for WBI relative to CI should disappear when we control for the experimental design. For these analyses, we proposed the following research questions:

Q6: Are WBI and CI equally effective with respect to participant learning when the same instructional methods are used to deliver training?

Q7: Are WBI and CI equally effective with respect to participant reactions when the same instructional methods are used to deliver training?

Q8: Does the experimental design moderate learning from WBI relative to CI? That is, will any effects observed for all studies also be evident when examining only studies with true experimental designs?

Q9: Does the experimental design moderate reactions to WBI relative to CI? That is, will any effects observed for all studies also be evident when examining only studies with true experimental designs?

Learner control. The final set of research questions address two specific aspects of the learning environment: the level of learner control and the length of training. Learner control refers to the extent to which trainees have control over their learning experience by affecting the content, sequence, or pace of material (Friend & Cole, 1990). The absence of learner control is characterized by program control in which the instructional software controls most or all of the decisions in an online learning context.

A purported advantage of WBI is it typically provides trainees with more control than traditional CI (Welsh, Wanberg, Brown, & Simmering, 2003). Adults tend to believe they know what they need to learn and know how much time they must spend studying (Knowles, 1990).
While many individual studies have reported that adults react favorably to receiving control during instruction (e.g., Andriole, 1995; Park & Tennyson, 1983; Tennyson & Buttrey, 1980), meta-analytic investigations have reported no consistent overall effects for learner control on learner reactions (Kraiger & Jerden, in press; Niemiec, Sikorski, & Walberg, 1996).

Accordingly, we were curious as to whether learner control would moderate reactions to WBI relative to CI such that learners would react more positively when WBI was rated high or low in learner control while holding the level of control constant in CI.

Research also shows that the impact of learner control on actual learning is negligible (Kraiger & Jerden, in press; Niemiec et al., 1996). Since prior research has not consistently demonstrated an effect for learner control on learner achievement, we cannot develop a directional hypothesis regarding the moderating effect of learner control on the effectiveness of WBI. However, given the great potential for individual customization in online courses, we were interested in the impact of learner control during WBI. Thus, we proposed the following research questions:

\[ Q10: \text{Will the level of learner control moderate participant learning in WBI relative to CI? Relative to CI, will participants learn more with low or high levels of learner control in WBI environments?} \]

\[ Q11: \text{Will the level of learner control moderate reactions to WBI relative to CI. Relative to CI, will participants react more favorably towards low or high levels of learner-controlled in WBI environments?} \]

\[ Length \text{ of training. We also explored the effects of the length of training on learning from and reactions towards WBI relative to CI. The training programs we reviewed varied tremendously in length. Length of the training courses ranged from 1-120 days. Course length} \]
may differentially influence learning or reactions depending on whether trainees become more proficient at learning or whether early novelty effects wear off. Thus, we were curious as to whether the effectiveness of WBI relative to CI decreased, increased, or remained the same as the length of the course increased. Thus, we proposed two additional research questions relating training characteristics to participants’ learning and reactions:

**Q12:** Does the length of training moderate learning from WBI relative to CI? Relative to CI, will participants learn more, less, or the same amount from WBI as course length increases?

**Q13:** Does the length of training moderate reactions to WBI relative to CI? Relative to CI, will participants react more favorably, less favorably, or the same towards WBI as course length increases?

**Method**

**Literature Search**

A computer-based literature search of PsycInfo and ERIC was used to locate studies in the training and education literature from 1996 to February 2005. Since the technology for online instruction is relatively new, we designated 1996 as a reasonable cutoff date for evaluations of WBI. We scanned references of the obtained studies for earlier citations and found only two relevant studies published prior to 1996.

In order to be included in the initial review of abstracts, each abstract had to contain a term relevant to the Internet and reactions or learning outcomes. To meet the search criteria some combination of the keywords: *Web, online, or Internet and evaluate, learn, transfer, behavior, performance, knowledge, satisfaction, dissatisfaction, reaction, achieve, or outcome* had to be present. The initial computer search resulted in a list of 3,461 possible studies. A
review of titles and abstracts reduced the list to 249 studies potentially containing relevant information. A careful reading of this list resulted in 59 articles included in the current study. The electronic search was supplemented with manual searches of the reference lists from Allen, Bourhis, Burrell, and Mabry (2002), Bernard et al. (2004), Hsu (2003), Olson and Wisher (2002), and Paul (2001), as well as a manual search of the *Journal of Asynchronous Learning Networks* from 1996 to 2004. Manual searches contributed an additional 33 studies to the current review.

We also searched for unpublished studies. First, a request was sent to the Advanced Distributed Learning listserv of over 8,000 people working in the area of training and development. Second, authors of annual review chapters on training (Campbell, 1971; Goldstein, 1980; Latham, 1988; Salas & Cannon-Bowers, 2001; Tannenbaum & Yukl, 1992; Wexley, 1984) and training textbooks (Blanchard & Thacker, 2004; Goldstein & Ford, 2002; Noe, 2005; Saks & Haccoun, 2004; Wexley & Latham, 2002) were asked to provide leads to unpublished work, as well as any manuscripts they may have. Third, consultants who listed training evaluation as an area of expertise on the Society of Industrial-Organizational Psychology (SIOP) Consultant Locator (http://www.siop.org/sioplocator) were contacted via e-mail. Fourth, the SIOP and Academy of Management conference programs from 1996-2005 were manually searched to locate relevant studies. These efforts identified an additional four studies, yielding a total of 96 studies that met the criteria for inclusion in the current review.

**Inclusion Criteria**

The goal of the literature search was to identify all research reports where college students or employees were acquiring knowledge or skills to prepare them for current or future employment opportunities. Initially we gathered research reports that reported gain scores from
participating in WBI or that compared learning from or reactions following WBI and CI or WBI-S and CI. However, due to the upward bias in effect sizes from gain score research (Lipsey & Wilson, 2001), the current report focuses exclusively on studies that compared the effectiveness of WBI or WBI-S to CI. WBI was defined as a course where the material is delivered via the Internet. Trainees do not meet with the instructor face-to-face with the exception of a possible initial orientation or to complete exams throughout the course. CI was defined as a course where the material is delivered face-to-face via an instructor. WBI-S is a course that delivers material via the Internet and face-to-face via an instructor.

Studies had to meet five criteria to be included in the current review: 1) the study compared the effectiveness of WBI or WBI-S to CI for delivering material on the same topic; 2) the article was written in English; 3) the article reported results that allowed the calculation of a $d$ statistic (e.g., group means and standard deviations, a t-test, or univariate F-test) or the author complied with a request to provide this information; 4) study participants were non-disabled adults age 18 or older; 5) training was conducted on a topic that provides job-related knowledge or skills. The last two criteria were used to support generalization to a population of adults participating in workplace training.

**Data Set**

*Nonindependence.* Decisions about non-independent data points (i.e., multiple effect sizes from one sample) should take into account whether the effect sizes assess similar or different constructs (Arthur, Bennett, & Huffcutt, 2001). The criterion types of interest include trainee reactions, declarative knowledge, and procedural knowledge. Effect sizes calculated for different types of criteria were considered to be independent and retained as separate data points even if they were from the same sample. Occasionally a single study would report data from two
Web-based training groups and/or two classroom groups. In these situations, an effect size was calculated for all possible Web-classroom pairs and averaged by weighting each of the effect sizes by the sum of the sample size of the independent training group and one half of the sample size of the nonindependent group. Thus, the nonindependent sample was weighted according to its sample size in the overall effect size. In addition, 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 in the meta-analysis. For example, most of the effect sizes for declarative knowledge were based on data collected immediately after training.

Coding and Interrater Agreement

In addition to recording all relevant effect sizes, samples sizes, and reliabilities, the following information was coded from each study: a) reaction measures, b) training outcome criteria (i.e., declarative and/or procedural knowledge), c) similarity of instructional methods in WBI and CI, d) experimental design, e) degree of learner control, and f) length of training. Coding rules are described below. Scales for each moderator were drafted prior to coding and modified following initial attempts to code articles and reach a consensus.

Training reactions. We initially sought to code and investigate the comparative effectiveness of WBI and CI on specific dimensions of training reactions (e.g., affect v. utility). However, there was insufficient detail in many studies to code for specific reactions dimensions and too few studies within certain reactions dimensions. Accordingly, while separate dimensions of training reactions were coded, specific dimensions were treated as indicators of an overall satisfaction construct by aggregating all studies that reported any reaction effect size in a single analysis. To avoid violating the assumption of independence, when multiple reactions were
reported in a single study, the effect sizes were averaged.

**Learning outcomes.** Declarative and procedural knowledge were coded based on the Kraiger et al. (1993) multi-dimensional framework of learning. Declarative outcomes were defined as cognitive and structural knowledge assessments designed to assess if the trainees remembered concepts presented in training; they were always assessed with a written test. Procedural outcomes were defined as the ability to perform the skills taught in training. They were assessed by participating in an activity (e.g., simulations or role plays) or written test that required trainees to demonstrate memory of the steps required to complete the skills taught in training. For example, Browning (1999) taught an undergraduate course on educational technology and evaluated the course with both declarative knowledge and procedural knowledge assessments. The declarative knowledge assessment consisted of a multiple choice and fill-in-the-blank exam designed to assess understanding of the concepts taught in the course. The procedural knowledge assessment required trainees to perform the software application skills taught in training.

**Similarity of instructional methods.** Similarity of instructional methods across media was coded on a two-point scale. An instructional method is a technique used to deliver training content (e.g., lecture, online tutorials, video, textbooks). WBI and CI had similar methods when all of the methods present in WBI had comparable methods present in CI. An example is a management information systems course researched by Carey (2001). Both WBI and CI included a textbook, practice exams, and assignments. CI received lecture and discussed with the instructor face-to-face while WBI received a copy of the PowerPoint slides from the lecture online and e-mailed with the instructor. WBI and CI had different instructional methods whenever a method was present in one form of instruction and there was not a comparable
method in the other form of instruction. An example is an introductory psychology course researched by Taylor (2002). In this instance CI was delivered via lecture, quizzes, and a textbook while WBI was delivered via textbook, quizzes, assignments, discussion board, peer evaluation, and e-mail.

Experimental design. Research reports utilized an experimental design when trainees were randomly assigned to WBI and CI. Research reports utilized a quasi-experimental design when trainees self-selected into WBI or CI.

Learner control. Learner control can include control over the content, sequence, and pace (Friend & Cole, 1990). We originally explored ways of coding multiple levels of control within dimensions, but found that there was insufficient detail in most research reports to do so. Thus, learner control was coded on a two-point scale separating little or no control (hereafter, low) from moderate or high levels of control (hereafter, high). In the present study, learner control was coded as being low if trainees had little or no control over the content, sequence or pace. An example of a Web-based course with little or no control is a non-interactive lecture-based class. Learner control was high when trainees had at least some control over two of the three dimensions—pace, content, or sequence. An example of a course with a high level of control is a managerial course where trainees can select material that is relevant to their jobs.

Length of training. Length of training was coded as the number of days trainees spent in WBI and CI.

Coding Agreement. All articles were coded independently by two trained raters. The initial mean level of agreement across all of the studies coded was 91%. The two coders then discussed discrepancies and came to a consensus. After discussing all discrepancies, 100% agreement was reached.
Calculating Effect Size Statistic (d) and Analyses

The Hedges and Olkin (1985) approach was used to analyze the data. The effect size calculated for each study was d, the difference between the Web and classroom training groups, divided by the pooled standard deviation. When means and standard deviations were not available, effect sizes were calculated from a t-test or univariate F-test based on the formulas reported in Glass, McGaw, and Smith (1981) and Hunter and Schmidt (1990).

Effect sizes were corrected for small sample bias using the formulas provided by Hedges and Olkin (1985). We then corrected the reactions effect sizes for attenuation using the scale reliabilities reported in each study. When a study failed to provide a coefficient alpha reliability estimate, we used the average reliability for the variable across all samples from the current study and from the Sitzmann, Casper, Brown, Witzberger and Polliard (2003) meta-analysis. While we aggregated all effect sizes for reaction measures, we corrected effect sizes at the study level based on the type of reaction measure. The average reliabilities were .83 for measures of affective, utility and difficulty reactions, .87 for instructor reactions, .79 for delivery reactions, and .84 for general reactions. We did not correct the declarative or procedural knowledge effect sizes for attenuation due to the lack of availability of test-retest or alternate forms reliability coefficients. Finally, 95% confidence intervals were calculated around the weighted mean ds. Confidence intervals assess the accuracy of the estimate of the mean effect size and provide an estimate of the extent to which sampling error remains in the weighted mean effect size (Whitener, 1990).

Outliers Analysis

We computed Huffcutt and Arthur’s (1995) sample-adjusted meta-analytic deviancy (SAMD) statistic to identify outliers. This procedure identified one declarative knowledge
outlier reported by Vessell (2000). Students in CI accessed course material that was intended to be exclusively utilized by students in WBI-S, providing students in CI with a competitive advantage and resulting in CI outperforming WBI-S. The associated SAMD value of 10.8 was more than twice the value of the next data point. In addition, one reaction outlier reported by Stadtlander (1998) was identified in which students in WBI encountered extensive technical difficulties, resulting in CI outperforming WBI. The associated SAMD value of 10.76 was more than twice the value of the next data point. All of the analyses were run with and without the outliers. The results of the two sets of analyses were virtually identical. Thus, only the results with outliers removed are included in the current manuscript.

Moderator Analysis

Hedges and Olkin’s (1985) homogeneity analysis was used to determine whether the effect sizes were consistent across studies. For the main effect analyses, the set of effect sizes was tested for homogeneity with the $Q_T$ statistic. $Q_T$ has an approximate $\chi^2$ distribution with $k - 1$ degrees of freedom, where $k$ is the number of effect sizes. If $Q_T$ exceeds the critical value, then the null hypothesis of homogeneity is rejected. Rejection indicates there is more variability in the effect sizes than expected by chance fluctuations, identifying the potential for moderators.

The goal of the moderator analysis was to focus exclusively on studies that were consistent in their operationalization of WBI and CI. Oswald & McCloy (2003) recommend narrowing the set of studies included in meta-analyses to a subset of studies that are theoretically and rationally similar to each other. Thus, we chose to eliminate studies where the Web was used as a supplement to CI from the moderator analysis. The elimination of WBI-S studies reduced the analysis sample but increased the interpretability of the results. We also focused our learner control moderator analyses on classroom courses low in learner control. Only four
reports were based on CI that was high in learner control. We eliminated these studies from the learner control analyses to increase the interpretability of the results. This allowed us to compare effect sizes between WBI low in learner control to WBI higher in learner control. In the training length moderator analyses, we focused on studies where the number of days spent in training was the same for WBI and CI (eliminating four studies). This allowed us to examine the effect of varying course length on the relative effectiveness of WBI and CI.

The moderating effects of categorical variables were tested by classifying studies according to the moderator categories and testing for homogeneity between and within categories (Lipsey & Wilson, 2001). For each categorical moderator, a between-class goodness-of-fit statistic, $Q_B$, was calculated to test for homogeneity of effect sizes across moderator categories. It has an approximate $\chi^2$ distribution with $p - 1$ degrees of freedom, where $p$ is the number of moderator categories. If $Q_B$ exceeds the critical value, it indicates a significant difference across the moderator categories and is analogous to a significant main effect in ANOVA. In addition, a within-class goodness-of-fit statistic, $Q_w$, was calculated to test for homogeneity of effect sizes within each moderator category. It has an approximate $\chi^2$ distribution with $m - 1$ degrees of freedom, where $m$ is the number of effect sizes across all of the moderator categories. If $Q_w$ exceeds the critical value, it indicates the effect sizes within the moderator categories are heterogeneous. Finally, the moderating effect of length of training was tested with inverse variance weighted correlations between the moderator variable and the effect sizes, for attenuation in the case of reactions.

Results

Ninety-six research reports contributed data to the current meta-analysis, including 65 published studies, 18 dissertations, and 13 unpublished studies. These reports reported data
gathered from 19,331 people who took part in 168 training courses. The topic of the training courses varied greatly and included psychology, engineering, computer programming, business, and technical writing courses. In 67% of research reports the trainees were undergraduates, while trainees were graduate students (18% of courses) or employees (15% of courses) in the remaining studies. Across all studies providing demographic information, the average age of participants was 24 years and 41% of the participants were male.

**Relative Effectiveness of WBI and WBI-S**

The first and second research questions addressed the effectiveness of WBI relative to CI on the acquisition of declarative and procedural knowledge. As shown in Table 1, across all studies, the declarative knowledge effect size was .15, indicating that on average WBI is 6% more effective than CI for teaching declarative knowledge. Moreover, the confidence interval for effects on declarative knowledge excluded zero. The WBI procedural knowledge effect size was near zero \( (d = -.07) \) and the confidence interval contained zero, suggesting WBI and CI are equally effective for teaching procedural knowledge. Thus, across all studies, there is evidence that WBI is more effective than CI for teaching declarative knowledge, but not for teaching procedural knowledge.

---

**Insert Table 1 about Here**

---

We also examined whether differences in learning outcomes from WBI relative to CI were greater when the Web was the sole means of instruction or when the Web was used as a supplement to CI, addressing research question three. Both the declarative and procedural knowledge effect sizes were larger for WBI-S than WBI. The WBI-S effect size was .34 for
declarative knowledge and .52 for procedural knowledge suggesting WBI-S is 13% more effective than CI for teaching declarative knowledge and 20% more effective than CI for teaching procedural knowledge. The 95% confidence intervals for both declarative and procedural knowledge excluded zero. Thus, CI supplemented by WBI was more effective than CI alone for teaching declarative and procedural knowledge. Together, the effect sizes in Table 1 indicate that, WBI was more effective than CI for training declarative knowledge (but not procedural knowledge). Additionally, WBI-S was more effective than CI for training declarative and procedural knowledge.

The fourth and fifth research questions address reactions towards WBI and WBI-S relative to CI. When comparing WBI to CI, the mean corrected effect size was zero, suggesting trainees are equally satisfied with the two delivery media. However, the mean corrected effect size for the WBI-S vs. CI comparison was negative ($d = -0.15$) and the 95% confidence interval excluded zero. Trainees reacted 6% more favorably towards CI than WBI-S.

The $Q_T$ statistic for all six effect sizes reported in Table 1 were statistically significant, suggesting there are potential moderators of the effectiveness of WBI and WBI-S relative to CI. While we looked for main effects for each type of instruction for reactions, procedural knowledge, and declarative knowledge, the moderator analyses will focus exclusively on declarative knowledge and reactions outcomes comparing WBI to CI. We do so for two reasons. First, only a few studies examined the effectiveness of WBI and WBI-S relative to CI for teaching procedural knowledge. Thus, insufficient data was available to examine moderators of the relative effectiveness of the delivery media for teaching procedural knowledge. Second, mixed training delivery methods used in WBI-S make it difficult to interpret moderator analyses. That is, if a moderator is having an effect, it is difficult to assess if the effect is due to
characteristics of WBI, CI, or both components of the course. Thus, WBI-S data was not
included in the moderator analyses. Conducting focused analyses will allow us to draw stronger
conclusions regarding moderators of the effectiveness of WBI relative to CI.

**Moderator Analyses**

The next set of research questions examined moderators of the effectiveness of WBI
relative to CI. Table 2 shows mean effect sizes and estimates of homogeneity within moderator
subgroups ($Q_B$). A significant $Q_B$ indicates the mean effect sizes across categories of the
moderator variable differ by more than sampling error (Lipsey & Wilson, 2001). That is, the
moderator variable is having an effect. The $Q_B$ statistic was significant for all four moderators for
declarative knowledge but only similarity of instructional methods was a significant moderator
of reactions.

-----------------------------------------
Insert Table 2 about Here
-----------------------------------------

**Experimental Characteristics.** The sixth and seventh research questions addressed the
relative effectiveness of WBI and CI when the instructional methods were the same across
delivery media. The declarative knowledge effect size was near zero when the same
instructional methods were used to deliver WBI and CI ($d = .04$). However, WBI was 11% more
effective than CI for teaching declarative knowledge when different instructional methods were
used to deliver the two courses ($d = .29$). This pattern of results supports Clark’s (1983, 1994)
theory that differences in instructional methods are more influential than differences in delivery
media in their impact on learning. In addition, trainees reacted 7% more favorably towards CI
than WBI when the same methods were used to deliver instruction ($d = -.17$) and 7% more
favorably towards WBI than CI when different methods were used to deliver instruction ($d = .17$). Thus, when WBI and CI use the same methods, learners on average prefer CI.

The eighth and ninth research questions addressed effects of the research design on study outcomes. While we found a positive mean effect size for declarative knowledge in quasi-experimental studies ($d = .18$), CI was 10% more effective than WBI for teaching declarative knowledge when trainees were randomly assigned to courses ($d = -.26$). Experimental design did not moderate reactions towards WBI relative to CI ($d = -.02, .01$ for experimental and quasi-experimental designs, respectively).

Training Design Characteristics. Research questions 10 and 11 addressed the impact of training design characteristics on the acquisition of declarative knowledge from and reactions towards WBI relative to CI. Note that in both analyses, the level of learner control was low in all of the classroom courses, allowing us to examine the effect of varying levels of learner control in WBI on training outcomes. The results indicated that the level of learner control moderated the acquisition of declarative knowledge from WBI compared to CI. WBI trainees learned more than CI trainees when they were afforded a high level of control ($d = .30$) than when they were afforded little control ($d = .07$). The level of learner control did not affect reactions towards WBI relative to CI ($d = -.01, -.01$ for low and high learner control, respectively).

The twelfth and thirteenth research questions addressed the effect of length of training on learning in and reactions towards WBI relative to CI. Note that in all of the courses included in the analyses, the number of instructional days was the same for WBI and CI. Two analysis strategies were used in these investigations. First, length of the training course was analyzed as a dichotomous variable to compare the strength of the moderator effect to the three previous moderator results (See Table 2). Length of training was dichotomized such that short courses
were categorized as those spanning less than 80 days and long courses spanned 80 or more days. In long courses, the mean effect size was positive \( (d = .17) \); in short training courses, the mean effect size was negative \( (d = -.18) \). Differences in training length did not strongly influence reactions to training \( (d = 1.14, -.02 \text{ for short and long courses, respectively}) \).

Second, inverse variance weighted correlations were used to assess the effect of the number of days of training on both learning in WBI and reactions towards WBI relative to CI. The number of days of training was positively and significantly correlated with the declarative knowledge effect size (weighted \( r = .38; p < .05 \)), indicating WBI trainees gain more declarative knowledge relative to CI as the length of the class increased. Length of training was uncorrelated with the reactions effect sizes (weighted \( r = -.01; p > .05 \)).

Overall the moderator results indicated that all four moderators had an effect on the acquisition of declarative knowledge from WBI relative to CI while similarity of training methods was the only moderator with a significant effect on reactions to training. However, for all of the declarative knowledge and reaction moderator results the \( Q_w \) was significant, indicating there is more variation within the moderator categories than would be expected by subject-level sampling error alone (Lipsey & Wilson, 2001). That is, none of the moderator variables can independently account for all of the variability in declarative knowledge or reactions effect sizes across studies.

Discussion

Meta-analytic procedures were used to examine the effectiveness of WBI and WBI-S compared to CI. The goals of the research were to determine whether WBI and WBI-S are as effective as CI for teaching declarative and procedural knowledge, whether trainees react more
favorably towards WBI, WBI-S, or CI, and whether there are experimental and training context variables that moderate the effects of WBI relative to CI on learning or learner reactions. We will discuss both the practical and theoretical implications of our results, as well as limitations of the study and directions for future research.

Across all studies, the results indicated that WBI was 6% more effective than CI for teaching declarative knowledge. These results are based on 71 effect sizes and 10,910 learners. WBI and CI were equally effective for teaching procedural knowledge and trainees were equally satisfied with the two delivery media.

The results were somewhat different when we examined instances of blended learning – online learning used to supplement face-to-face instruction (WBI-S). Across all studies, the results indicated that combining classroom and Internet components was more effective than stand alone CI for teaching trainees job-relevant knowledge and skills. WBI-S was 13% more effective than CI for teaching declarative knowledge and 20% more effective than CI for teaching procedural knowledge. Similar meta-analytic findings were reported by Zhao et al. (2005) who found that “mixed method” or blended distance courses result in better outcomes than distance education or face-to-face instruction alone. While there is yet no theory that explains why blended learning would more effective than WBI or CI alone, the results are consistent with recent calls for greater use of this instructional medium (e.g., Alavi & Leidner, 2001; Greco, 1999; Masie, 2002; Trierweller & Rivera, 2005). However, trainees react 6% more favorably towards stand alone CI than WBI-S. Thus, while converting to WBI-S from CI may improve learning, there may be a tradeoff in terms of trainee satisfaction.

Theoretical Implications

Advocates of WBI or TAI cite numerous potential pedagogical benefits including the use
of multi-media, learner customization, and opportunities for guided learning (Bailey & Cotlar, 1994; Dumont, 1996; Hiltz & Wellman, 1997; Liaw, 2001; Sullivan, 2001). However, other theorists argue that there is nothing uniquely advantageous to any instructional medium, so we should expect no effects in well-designed media comparison studies. This position is summarized by Clark (1983) who wrote that media are “mere vehicles used to deliver instruction but do not influence student achievement any more than the truck that delivers our groceries causes change in out nutrition” (p. 445). Thus, a secondary purpose of our study was to capitalize on the large number of studies analyzed and unique coding methods to investigate the veracity of Clark’s well-cited position.

Our results strongly support Clark’s position that media effects in single study research are largely spurious. We first note that across all studies, we found relatively small differences between WBI and CI on both measures of procedural and declarative knowledge (though confidence intervals for the latter outcome excluded zero). More importantly, we were able to examine the impact of the research design on study outcomes for declarative knowledge. We found that when trainees were randomly assigned to conditions, CI was more effective than WBI for teaching declarative knowledge \( (d = -.26) \). However, this result is in the opposite direction of the effect sizes for WBI relative to CI across all studies \( (d = .15) \) and across studies using a quasi-experimental design \( (d = .18) \). Thus, consistent with Clark’s arguments (1983; 1994; Clark & Sugrue, 1995), studies are more likely to provide support for WBI when research participants are allowed to self-select into courses.

The similarity of instructional methods moderator results added additional support for Clark’s theory. Clark argued that media comparison studies have confounded media with instructional methods, making it impossible to detect the true cause of differences in course
effectiveness. In the current meta-analysis, WBI and CI were equally effective for teaching declarative knowledge when similar instructional methods were used to deliver the two courses, supporting Clark’s theory. This suggests that instructional methods are driving differences in the effectiveness of WBI relative to CI. In addition, WBI was on average 11% more effective than CI when different methods were used to deliver the two courses. A qualitative analysis of research reports identified two characteristics of research reports where different training methods were used and trainees learned more from WBI than CI. First, the Internet courses tended to incorporate more instructional methods than CI. Utilizing a variety of instructional methods may allow trainees to tailor the course to be consistent with their learning styles (Salomon, 1988). Second, the Internet courses tended to require students to be more active than CI. This is consistent with Webster and Hackley’s (1997) guidelines for teaching in distance learning, “learning is best accomplished through the active involvement of the students” (p. 1284). Spending time practicing the key task components of training should help trainees develop an understanding of the deeper, structural features of the task (Newell, Rosenbloom, & Laird, 1989). Frequent practice should also increase the likelihood that trainees will automate skills by the end of training, leading to better performance at the end of training (Rogers, Maurer, Salas, & Fisk, 1997). Thus, it is critical that CI requires trainees to as active as WBI and incorporates as many instructional methods as WBI to promote similar learning outcomes between the two delivery media.

Together, our findings and observations suggest that instructional methods may be more important than delivery media for ensuring effective learning. While these results have important theoretical implications, they have practical implications as well. Practical implications of study findings will be addressed in the following section.
Practical Implications

The current meta-analytic results have several direct implications for organizations and institutions considering implementing online learning programs. Advocates of online instruction (e.g., Galagan, 2001; Goodridge, 2001; Hall, 1997) suggest that it can be a more cost-effective means of training than face-to-face instruction, although well-controlled studies documenting the cost-effectiveness or utility of WBI are rare (Welsh et al., 2003). Assuming that over time WBI is less expensive than CI, even findings that that show no mean differences between WBI and CI provide support for implementing online instruction. The results we report can be used in conjunction with accurate estimates of the cost of implementing and maintaining online instruction programs to estimate the utility (see Mathieu & Leonard, 1987) of converting face-to-face-training to online.

The results also indicate care should be taken whenever organizations and institutions consider implementing WBI as the relative effectiveness of the training may depend on both the intended learning outcomes and the training conditions. Given that WBI is at least as effective as CI for teaching job relevant knowledge and skills when trainees are allowed to self-select into courses, the current results can be used by organizations and universities to justify the expenditure necessary to develop online instruction. However, they must be cautious about completely replacing CI with WBI. Researchers are beginning to understand that face-to-face and online instruction create very different learning environments (Arbaugh, 2005; Dumont, 1996; LaRose & Whitten, 2000). Thus, forcing trainees into online courses may result in some trainees failing to master the course material. Accordingly, the moderator analyses we conducted are helpful for understanding conditions that influence the effectiveness of WBI. Our results indicated that learners acquired relatively more declarative knowledge from WBI than CI.
when different instructional methods were used, courses were longer, and learners were afforded more control over the instructional such as pace and content. We return to the issue of designing more effective Web-based courses below.

It is important to note that the positive effect size for declarative knowledge across all studies was reversed when trainees were randomly assigned to courses. There are several possible explanations for these findings. First, it is possible that trainees who are higher in motivation or cognitive ability are self-selecting into WBI when they are allowed to choose which course to sign up for. Thus, pre-existing differences between trainees who prefer WBI and trainees who prefer CI may result in the appearance that WBI is more effective than CI. Second, trainees who lack technical skills may be forced to participate in WBI when trainees are randomly assigned to courses. Providing trainees with a computer and Internet skills course before participating in WBI may result in the two delivery media being equally effective for teaching declarative knowledge. Third, the advantage of CI may be due to the value added by onsite instructors. This is consistent with the instructor immediacy literature which suggests instructors can use verbal and nonverbal communication strategies to motivate students, resulting in more positive reactions and greater learning (Christophel, 1990). Thus, additional research is needed to explore differences in the results of experimental and quasi-experimental studies. An experiment where half of the trainees are allowed to self-select into WBI and CI and the other half of participants are randomly assigned to courses would allow researchers to disentangle differences in the effectiveness of the delivery media for teaching declarative knowledge.

Designing More Effective Online Training Courses

The current study investigated the effect of two course design characteristics on the effectiveness of WBI relative to CI. Across studies, the extent to which Web-based trainees
learn more than classroom trainees is greatest when Web-based trainees are provided with control and in long courses. Under these conditions, the declarative knowledge effect size was .35, suggesting WBI was 14% more effective than CI. In contrast, it is also possible to design Web-based courses in which learning levels will be inferior to CI. CI was 16% more effective than WBI when the courses lasted less than 80 days and WBI failed to provide control to learners ($d = -.42$). Thus, attention to course design features is critical for maximizing learning outcomes.

As online learning may be a new experience for many trainees, longer training programs may give learners the opportunity to adapt to the technology. That is, trainees may learn how best to learn in these environments. One of the demonstrated advantages of WBI is the opportunity to develop collaborative learning communities (e.g., Alavi, Wheeler, & Valacich, 1995; Rovai, 2001), but it takes learners time to build and benefit from collaborative contexts (Duffy & Kirkley, 2004; Garrison, 2003). Accordingly, it would be interesting to test inexperienced participants at multiple occasions in a Web-based training course to determine whether they are using more adaptive learning strategies over time and how collaborative learning environments facilitate learning over time. Additionally, more research is needed to understand the effects of cohort size, peer-to-peer interactions, and synchronous v. asynchronous communication of the effectiveness of WBI.

We also found a moderating effect for learner control on declarative knowledge effect sizes. Compared to classroom learners, participants in WBI learned more when given a high level of learner control. Learner control may be provided along a number of dimensions such as content, sequence, or pace and research has suggested that various dimensions of learner control may differ in their effects on learning from WBI (Lunts, 1997). Due to limited descriptions of training courses in many research reports, we were unable to distinguish among the learner
control dimensions in our coding. Thus, future primary research should provide more detailed
descriptions of training courses to allow more precise coding and evaluation of learner control in
future meta-analyses. More research is also needed to understand which specific learner control
options online learners prefer and which facilitate learning.

**Comparison to Previous Technology-Assisted Instruction Meta-analyses**

It is worth noting that the overall positive effect size for WBI compared to CI is smaller
than those reported in meta-analyses of other types of TAI (Fletcher, 1990; Kulik, 1994; Kulik &
Kulik, 1991; Liao, 1999; Yaakub, 1998), although similar to a recent meta-analysis of distance
education (Zhao et al., 2005). There are several possible explanations for this. In contrast to
previous meta-analyses, ours used an adult population learning work-related knowledge and
skills. Also note that WBI is a relatively new training platform, and as such, its overall
effectiveness may be compromised by several non-permanent conditions. For example, in many
studies there may have been insufficient bandwidth to optimize training delivery or trainees may
have lacked the technical skills needed to access the instructional content (Welsh et al., 2003).
Over time, instructional designers may make more informed decisions about how to structure
Web-based environments to ensure greater learning. Accordingly, it is important to identify
variables that influence the effectiveness of WBI courses.

In addition, our meta-analysis contained more studies and more unpublished studies than
prior meta-analyses of TAI. It is possible that previous meta-analyses reflect a publication bias
or other sampling problems not evident in the larger number of studies we were able to locate
and code. Regardless, since WBI, video-disks, and single-work station computer-based training
are each options for training delivery, other researchers may want to explore possible differences
in the relative effectiveness of these different types of TAI.
**Study Limitations**

There are several limitations to the current study. While we would have preferred to investigate the impact of WBI relative to CI based on three categories of learning outcomes, we were able to identify only 12 studies that assessed procedural knowledge and even fewer studies that assessed affective learning. In the latter case, there was an insufficient number of studies to determine an overall effect size, and in the former case, there was an insufficient number of studies to examine potential moderators. Thus, we could not determine whether online learning is more or less effective for affective outcomes than the overall effect sizes reported for declarative and procedural knowledge. This is not merely an academic question; an increasing number of organizations are implementing WBI for diversity and sexual harassment training. In such programs, changing participants’ attitudes towards groups of employees is the desired outcome of these programs. Yet, little is known about the effectiveness of WBI in this regard. It is also possible that the size or direction of the moderating effects we found for declarative knowledge might differ if the learning outcome was procedural knowledge. Additional primary research is needed to examine the effectiveness of WBI for conveying affective and procedural knowledge.

Estimates of effect size heterogeneity within moderator categories also suggest there are moderators of the effectiveness of WBI that were not identified in the current study. Learners’ prior experience with WBI, course content, and course quality are potential moderators that we could not code due to lack of detail in research reports. Additional research is needed to examine the impact of these moderator variables on the effectiveness of WBI.

**Conclusion**

The current meta-analysis identified 96 studies reporting data from 19,331 trainees who
took part in 168 training courses. Across all of these reports, CI was more effective than WBI for teaching declarative knowledge when trainees were randomly assigned to courses and trainees were equally satisfied with the two delivery media. However, trainees learned the same amount from WBI and CI when the same instructional methods were used to deliver training. Overall these results strongly support Clark’s (1983, 1994) argument that instructional content and methods rather than delivery media determine learning outcomes. In addition, designing long training courses and providing trainees with control during WBI will maximize learning from WBI relative to CI.
Author Note

The authors are indebted to Richard Mayer, Kevin Ford, and Kenneth Brown for comments and suggestions on earlier versions of this manuscript.

This research was funded by the Department of Defense through contract number DASW01-03-C-0010. The views expressed here are those of the authors and do not necessarily reflect the views or policies of the Department of Defense.
Footnotes

1 Relative effectiveness refers to measures of WBI training effectiveness compared to CI. Because our meta-analysis compares WBI to CI, we cannot form any conclusions of the absolute effectiveness of WBI, but only in comparison to a traditional training method.

2 Results with outliers included in the analyses are available upon request from the first author.

3 A table with all of the hierarchical moderator results is available upon request from the first author.
References

(studies used in the meta-analysis are noted with an asterisk)


*Browning JB. (1999). *Analysis of concepts and skills acquisition differences between Web-delivered and classroom-delivered undergraduate instructional technology courses.*


*Caulfield JL. (2002). *Examining the effect of teaching method and learning style on work


for HEPE 129 – *Fitness and lifestyle management course*. Unpublished manuscript, Brigham Young University, Provo, Utah.


of Computing in Higher Education, 12, 57-83.

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


Communications of the ACM, 40, 44–49.

report on the implementation of courses offered on the Internet. Retrieved March 8, 1997,
from http://www2.ncsu.edu/ncsu/ltc/Project25/info/f97_assessment.html

Hsu Y. (2003). The effectiveness of computer-assisted instruction in statistics education: A meta-


Journal of Applied Psychology, 80, 327-334.

Hunter JE, Schmidt FL. (1990). Methods of meta-analysis: Correcting for bias in research

*Hurlburt RT. (2001). "Leclets" deliver content at a distance: Introductory statistics as a case

*Johnson M. (2002). Introductory biology online: Assessing outcomes of two student

satisfaction and learning outcomes in online and face-to-face learning environments.

Journal of Interactive Learning Research, 11, 29-49.


Web-Based Instruction 47


Mathieu JE, Leonard RL. (1987). An application of utility concepts to a supervisor skills training
program. *Academy of Management Journal*, 30, 316-335


*Mylona ZH. (1999).* Factors affecting enrollment satisfaction and persistence in a Web-based


*Sankaran SR, Sankaran D, Bui TX. (2000). Effect of student attitude to course format on


*Shafer ME, Lahner JM, Calderone WK, Davis JE, Petrie TA. (2002). The use and effectiveness


dissertation, University of California, San Diego.


*Vessell DC. (2000). Comparing the WWW and WebCT to traditional methods of supporting an
Web-Based Instruction 56


Whitener EM. (1990). Confusion of confidence intervals and credibility intervals in meta-


Table 1.

*Meta-analytic results for learning outcomes and reactions comparing Web-based instruction and Web supplements to classroom instruction.*

<table>
<thead>
<tr>
<th></th>
<th>d</th>
<th>Standard Error</th>
<th>k</th>
<th>N</th>
<th>Lower</th>
<th>Upper</th>
<th>Q_T</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Declarative knowledge</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WBI v. CI</td>
<td>.15</td>
<td>.02</td>
<td>71</td>
<td>10,910</td>
<td>.11</td>
<td>.19</td>
<td>267.49*</td>
</tr>
<tr>
<td>WBI-S v. CI</td>
<td>.34</td>
<td>.03</td>
<td>33</td>
<td>6,799</td>
<td>.29</td>
<td>.39</td>
<td>135.26*</td>
</tr>
<tr>
<td><strong>Procedural knowledge</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WBI v. CI</td>
<td>-.07</td>
<td>.07</td>
<td>12</td>
<td>944</td>
<td>-.20</td>
<td>.06</td>
<td>61.15*</td>
</tr>
<tr>
<td>WBI-S v. CI</td>
<td>.52</td>
<td>.09</td>
<td>6</td>
<td>507</td>
<td>.34</td>
<td>.70</td>
<td>23.33*</td>
</tr>
<tr>
<td><strong>Reactions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WBI v. CI</td>
<td>.00</td>
<td>.05</td>
<td>22</td>
<td>2,580</td>
<td>-.09</td>
<td>.09</td>
<td>51.78*</td>
</tr>
<tr>
<td>WBI-S v. CI</td>
<td>-.15</td>
<td>.06</td>
<td>11</td>
<td>1,769</td>
<td>-.26</td>
<td>-.05</td>
<td>119.67*</td>
</tr>
</tbody>
</table>

Notes. WBI = Web-based instruction; CI = classroom instruction; WBI-S = Web supplement to classroom instruction; $d =$ is the inverse variance weighted mean effect size; $k =$ the number of studies providing information included in the analysis; $N =$ sum of the sample sizes of studies providing information included in the analysis; $Q_T =$ homogeneity statistic

* indicates the $Q_T$ value is statistically significant at the .05 level and the effect sizes are heterogeneous
### Table 2.

**Meta-analytic moderator results comparing Web-based instruction to classroom instruction.**

<table>
<thead>
<tr>
<th></th>
<th>d</th>
<th>Standard Error</th>
<th>k</th>
<th>N</th>
<th>Lower</th>
<th>Upper</th>
<th>Q_B</th>
<th>Qw</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Declarative knowledge</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same methods</td>
<td>.04</td>
<td>.05</td>
<td>16</td>
<td>2,032</td>
<td>-.06</td>
<td>.13</td>
<td>17.43*</td>
<td>215.12*</td>
</tr>
<tr>
<td>Different methods</td>
<td>.29</td>
<td>.04</td>
<td>37</td>
<td>3,689</td>
<td>.22</td>
<td>.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>-.26</td>
<td>.09</td>
<td>11</td>
<td>529</td>
<td>-.43</td>
<td>-.08</td>
<td>22.96*</td>
<td>244.53*</td>
</tr>
<tr>
<td>Quasi-experimental</td>
<td>.18</td>
<td>.02</td>
<td>60</td>
<td>10,381</td>
<td>.13</td>
<td>.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low learner control</td>
<td>.07</td>
<td>.04</td>
<td>31</td>
<td>2,721</td>
<td>-.01</td>
<td>.15</td>
<td>15.13*</td>
<td>227.07*</td>
</tr>
<tr>
<td>High learner control</td>
<td>.30</td>
<td>.04</td>
<td>25</td>
<td>3,304</td>
<td>.22</td>
<td>.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short</td>
<td>-.18</td>
<td>.08</td>
<td>12</td>
<td>771</td>
<td>-.33</td>
<td>-.03</td>
<td>20.07*</td>
<td>215.26*</td>
</tr>
<tr>
<td>Long</td>
<td>.17</td>
<td>.02</td>
<td>53</td>
<td>8,796</td>
<td>.13</td>
<td>.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Different methods,</td>
<td>.40</td>
<td>.06</td>
<td>10</td>
<td>1,415</td>
<td>.29</td>
<td>.52</td>
<td>90.11*</td>
<td>113.02*</td>
</tr>
<tr>
<td>quasi-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>experimental,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>high control,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>long</td>
<td>-.79</td>
<td>.33</td>
<td>2</td>
<td>40</td>
<td>-1.43</td>
<td>-.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reactions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same methods</td>
<td>-.17</td>
<td>.07</td>
<td>6</td>
<td>1,190</td>
<td>-.30</td>
<td>-.03</td>
<td>10.83*</td>
<td>36.36*</td>
</tr>
<tr>
<td>Different methods</td>
<td>.17</td>
<td>.07</td>
<td>14</td>
<td>997</td>
<td>.02</td>
<td>.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>-.02</td>
<td>.14</td>
<td>5</td>
<td>255</td>
<td>-.29</td>
<td>.26</td>
<td>.02</td>
<td>51.76*</td>
</tr>
<tr>
<td>Quasi-experimental</td>
<td>.01</td>
<td>.05</td>
<td>17</td>
<td>2,325</td>
<td>-.09</td>
<td>.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low learner control</td>
<td>-.01</td>
<td>.07</td>
<td>10</td>
<td>1,260</td>
<td>-.15</td>
<td>.12</td>
<td>.00</td>
<td>47.19*</td>
</tr>
<tr>
<td>High learner control</td>
<td>-.01</td>
<td>.07</td>
<td>10</td>
<td>927</td>
<td>-.16</td>
<td>.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short</td>
<td>.14</td>
<td>.14</td>
<td>3</td>
<td>256</td>
<td>-.13</td>
<td>.41</td>
<td>1.23</td>
<td>50.05*</td>
</tr>
<tr>
<td>Long</td>
<td>-.02</td>
<td>.05</td>
<td>18</td>
<td>2,233</td>
<td>-.12</td>
<td>.08</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes.**  

*d* = is the inverse variance weighted mean effect size; *k* = the number of studies providing information included in the analysis; *N* = sum of the sample sizes of studies providing information included in the analysis; *Q_B* = between-class goodness-of-fit statistic; *Q_w* = within-
class goodness-of-fit statistic.

* indicates the Q value is statistically significant at the .05 level