8-2011

Best instructional practices for distance education: A meta-analysis

Robin Michael Roberts
University of Nevada, Las Vegas

BEST INSTRUCTIONAL PRACTICES FOR DISTANCE EDUCATION:

A META-ANALYSIS

by

Robin Michael Roberts

Bachelor of Arts
California State University, Fresno
1978

Master of Arts
Mennonite Brethren Biblical Seminary
1981

Master of Arts
California State University, Fresno
2002

A dissertation submitted in partial fulfillment
of the requirements for the

Doctor of Philosophy Degree in Learning and Technology
Department of Educational Psychology
College of Education

Graduate College
University of Nevada, Las Vegas
August 2011

THE GRADUATE COLLEGE

We recommend the dissertation prepared under our supervision by

Robin Michael Roberts

entitled

Best Instrumental Practices for Distance Education: A Meta-Analysis

be accepted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Learning & Technology
Department of Education Psychology

Neal Strudler, Committee Chair

Kendall Hartley, Committee Member

Gregory Schraw, Committee Member

James R. Crawford, Graduate Faculty Representative

Ronald Smith, Ph. D., Vice President for Research and Graduate Studies
and Dean of the Graduate College

May 2011

ABSTRACT

Best Instructional Practices for Distance Education:
A Meta-Analysis

By

Robin Michael Roberts

Dr. Neal Strudler, Dissertation Committee Chair
Professor of Curriculum and Instruction
University of Nevada, Las Vegas

Recent meta-analyses on the efficacy of distance education have concluded that no significant difference exists between face-to-face and distance education. At the same time, these meta-analyses noted that considerable heterogeneity existed between the individual studies used in the meta-analyses. Investigation of moderators responsible for that heterogeneity suggested that four things other than media delivery were primarily responsible for the majority of variation between study outcomes: methodological quality, instructor involvement, type of interaction, instructional methods and time-on-task. A comparative meta-analysis was performed to further investigate these moderators. Methodological quality, maturational differences in students and any undetermined media effects were controlled for through the inclusion process: Only Web-based courses delivered entirely at a distance (no blended courses were included) to adult learners and studies that were quasi-experimental or experimental in design were included. The effect of time-on-task on student outcomes is well documented in the literature and not addressed in the present study. A main effect for Web-based, adult distance instruction (g = .777; k = 59; SE = .078) was found. Results suggest Web-based distance education appears to have improved over time and that independent study, Behaviorist instructional strategies, instructor moderated collaboration, provision of formative feedback and the use of multimedia are more effective practices to use in Web-based distance education with adults. The need for more research into specific instructional strategies used in Web-based distance education and appropriate assessments for each is discussed.

ACKNOWLEDGEMENTS

A number of individuals provided invaluable support during the preparation of this dissertation and I would like to take this opportunity to thank them. First, my program and dissertation chair Dr. Neal Strudler stuck with me and kept me on track through those times when it seemed that this thing would never be completed. To him a special word of thanks is due and I still say that I made the right choice in choosing him as my chair so many years ago. The members of my committee also deserve thanks for accommodating my changes of mind and for sticking it out the process with me. A special word of thanks is also due Dr. Gregg Schraw, who provided special encouragement to this "burn child" when he most needed it. Dr. Kendall Hartley was gracious and generous enough to agree to serve on my committee when I sought him out, though neither of us thought it would take quite so long when we sat together outside the Lied library to discuss the possibility. Dr. James Crawford deserves special thanks for taking-up the mantel of outside advocate as a replacement for Dr. Patti Chance, who moved on to bigger and better things at California State University, San Diego.

It was through serendipity alone that I was privileged to work with Patti Chance in the Department of Educational Leadership. She provided support when I needed it and was as much a friend as she was a mentor. She, more than any other faculty member, deserves credit for getting me through the dark times and helping me to celebrate the good. I consider it great good fortune that I was assigned as Graduate Assistant to Patti—and thus to the Department of Educational Leadership—and I would like to thank everyone in that Department for being the bright light that outshone even the famed Las

Vegas strip; I am truly grateful for having been a part of your family and grieve at the impending dissolution of that Department—a casualty of the bad financial times.

Also worthy of mention are my fellow classmates in the Learning and Technology program, especially Fred Kuch and Mike McCreery. Mike was one of the first fellow students I got to know at UNLV and his cheerful enthusiasm and careful scholarship make him a pleasure to be around. I owe Fred Kuch more than I can express and consider him one of my closest friends. Fred generously made his home available for me to stay in during my trips to Vegas and is an academic research partner par excellence! We make a great team and together: watch-out, world! Thanks, Fred for everything.

Finally, none of this would be possible without the support of my family, my friends and my community, all of whom gave up something so I could pursue my dream. My good buddy, Dr. Mark Cave unselfishly let me off the hook for the many activities we have always done together for the duration of my Las Vegas stay. My good friend Kerry Arroues, provided encouragement with his usual sunny optimism and my close friend and band mate, Paul DaLuz commiserated through the final days of the dissertation preparation. My wife, Sylvia, kept the home fires burning and ran everything during my absence and also provided the necessary support—particularly when I was near to giving-up—that helped me finish. My daughters Terra and Tamara and their husbands Jeremy and Tom deserve major credit for my finally pursuing a doctorate. It was their marriages that provided the impetus for me to quit my teaching job and go after my Ph.D. My five grandchildren (and the one on the way!) all came into the family while I was working on my doctoral program and have made my retirement much more than I could ever have hoped for. Lastly, I am particularly pleased to provide my parents, Joe and Maureen

Roberts, with the ability to talk about their son, the doctor. They had to wait a long time for that to happen—I hope it is worth it.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

In the recent past, educational researchers have examined whether—and to what extent—digital computing technologies, especially personal computers, can be viable instructional tools (Charp, 2002; Mayers & Swafford, 1998; Ulmer, 1995). While that historical debate had not been completely been settled by the turn of the Twenty-first Century (*cf.* Clark, 2000; Cuban, 2002), to a great extent it has since been rendered superfluous: Computers have already become entrenched in American schools and it seems unlikely that they will disappear anytime in the near future (U.S. Dept. of Ed., 2004). Moreover, digital computing technologies have become a major part of the 21st Century American lifestyle and one likely to become even more ubiquitous (O'Reilly, 2000).

Closely associated with the educational use of digital computing technologies is the use of the Internet as an instructional tool and the rapid development of World Wide Web-based distance education. Web-based distance education has grown in recent years to the extent that by 2006, 66% of American colleges and universities offered Web-based courses (Parsad & Lewis, 2008). Some educators consider the growth in Web-based distance education to signal a paradigm shift in instruction (e.g., Desai, Hart & Richards, 2008).

The extent to which computers are responsible for these changes has been a matter for debate (Bohlin, 1997; Brown & Duguid, 2000; Cardwell, 1995; Colon & Simpson, 2003; Davis & Meyer, 1998; Oppenheimer, 1997; Patterson, 1996), but the presence of computers and the Internet in the classroom is no longer the question of the day; rather, how to leverage and use digital computing technologies to best effect has become the focus of most recent research (Bell, Schrum & Thompson, 2008).

Purpose of the Study

The purpose of this study is to identify effective instructional practices when distance education is the delivery method for higher educational instruction. This study seeks to provide a preliminary identification of the instructional practices and methods that appear to be more effective when used in a higher educational distance education setting by re-examining existing research using a type of research synthesis know as statistical meta-analysis. Specifically, the present study extends and “drills-down” into moderator factors previously identified in several recent, large-scale statistical meta-analyses using more focused lenses than have been used in the past.

The following research question guided this research: Which instructional practices are more effective when used in conjunction with Web-based, higher educational distance education? In a sense, this study explores potential best instructional practices for Web-based distance education used for higher education and provides a foundation for further, more detailed research on the subject.

Concomitant with this purpose is a brief secondary appraisal of the progress that has been made thus far in advancing DE instruction. In light of research reported over the past several decades, culminating in the meta-analyses examined here, it would be reasonable to expect some increase in the effectiveness of DE instruction during that time.

Accordingly, this study presents a brief trend analysis by comparing the aggregate effect of studies grouped in three chronological periods. The question of interest here is:

Has Web-based DE instruction benefited from the lessons provided by on-going research? A second research question that guides this study is: Have Web-based DE outcomes improved over time? The first research question requires that each study be coded according to the type of intervention used to create a contrast. The coding process for this study is briefly described below and in more detail in Chapter 2.

Background

Research Comparing Distance Education to Face-to-face Instruction

Much prior research has taken place examining how distance education (DE) compares with face-to-face instruction (f2f). Russell (1999) examined research comparing f2f and DE for the prior seventy years and found that the preponderance of evidence suggests that there is no significant difference between f2f learning and DE learning in either student attitudes or achievement—regardless of the medium employed for the delivery. Russell employed non-statistical methods in his study, leading some to criticize his conclusions on methodological grounds (e.g., Bernard et al., 2004a). However, statistical meta-analyses by Moore (1994) and Cavanaugh (2001) also reported similar, no significant difference findings between f2f and DE.

Four recent, large-scale statistical meta-analyses of studies comparing face-to-face and distance education support the historically consistent finding that no significant difference in student outcomes exists between DE and f2f courses: Bernard et al. (2004); Zhao, Lei, Yan, Lai and Tan (2005); Sitzmann, Kraiger, Stewart and Wisher (2009); and Means, Toyama, Murphy, Bakia and Jones (2009). All four studies concluded that *factors other than the media used to deliver instruction affected student outcomes*. In other words, the aggregate findings of research on the efficacy of distance education has consistently shown that, as asserted by Richard Clark (1983, 1994, 2000), media seems to be irrelevant—that is, there is no generic media effect on learning detectable by current methods of research.

Instead, Clark (1983, 1994, 2000) argued that the results of individual media comparison studies that indicated an advantage for one medium over another were confounded by differences in the instructional methods used, making it impossible to determine the true cause of differences in student outcomes between mediums. The four recent meta-analyses cited above statistically support what Clark earlier suspected: While the aggregate effect sizes for DE instruction compared to f2f instruction showed no significant difference in student outcomes, significant differences were found within each group. Specifically, individual studies of DE learning differed widely from each other and the same was true for the f2f portion of the studies that the meta-analyses examined (Bernard et al., 2004; Zhao et al., 2005; Sitzmann et al., 2005; Means et al., 2009). This effect was also noted in separate individual studies by Keefe (2003), Poirier and Feldman (2004) and Campbell et al. (2008).

The finding of no significant difference *across groups* and significant differences *within groups* indicates that some factor or factors—that is, confounds—other than the treatment are affecting the outcomes. When faced with such a statistical condition, the accepted practice is to attempt to identify the moderator factors that are confounding the findings. Each of the authors of the four aforementioned statistical meta-analyses conducted post-hoc statistical searches to identify factors that may have affected the student outcomes in the studies included in their meta-analyses.

Bernard et al. (2004) reported that the heterogeneity of the studies in their analysis was too great to identify any specific moderators but, using weighted multiple regression techniques, identified methodological quality and pedagogy as being significant sources of variance among studies. Zhao and colleagues (2005) found that instructor involvement, media involvement and the type of interaction were factors that moderated student outcomes in the studies included in their meta-analysis. Sitzmann et al. (2009) found that instructional methods were the source of differences in the effectiveness of the studies included in their meta-analysis. Finally, in the most exhaustive search for moderators conducted to date, Means and colleagues (2009) found that of the twenty-one factors they tested as potential moderators, only two emerged as statistically significant moderators of student achievement: time on task and equivalence of curriculum and instructional approach—that is, whether the instructional materials, learning activities and/or instructional resources used in the courses being compared were the same or different. Significantly, Means et al. did not further identify or differentiate what specific materials, activities or resources were examined in the sample studies included in their meta-analysis.

There seems to be agreement among these separate studies that media in and of itself has little discernable effect on student outcomes. Instead, the results of the recent meta-analyses discussed above that contrasted f2f and DE all agree that some aspect of research methodology or pedagogy explained a large part of the observed variance between studies within treatment groups (i.e., between f2f treatments and between DE treatments).

Research Comparing Distance Education Courses to other Distance Education Courses

In an effort to control for the effects of media delivery and methodological quality on student outcomes, Bernard et al. (2009) conducted a recent meta-analysis that compared DE courses to other DE courses in terms of the types of interactions that were afforded to students as part of those courses. Examining what they termed “interaction treatments,” Bernard et al. (2009) analyzed the effect of those interactions on student achievement. Interaction treatments, as used by Bernard et al., are intentionally planned and organized aspects of a course that foster, provide or afford for some type of interaction. They identified three types of interaction treatments that could be identified in DE courses: treatments that foster student-student interaction student-content interaction, and student-teacher interaction. They found that both student-student and student-content interaction treatments had more significant impacts on student achievement than did student-teacher interaction treatments and that student-student and student-content interaction treatments did not vary significantly from each other in their effects on student achievement. They also found that the greater the combined opportunities for interaction afforded during a course, the greater the effect of those interactions on student achievement. Bernard and colleagues (2009) concluded that student-content interaction treatments were the most effective of the three interaction treatments they studied for producing positive student achievement and suggested that “designing [Interaction Treatments] ITs into DE courses, whether to increase interaction with the material to be learned, with the course instructor, or with peers, positively affects student learning (p. 1264).”

Bernard et al. (2009) also concluded, as in the studies cited earlier, that there was a wide variability in the effect sizes between DE studies, and suggested that “fundamental confounds associated with different media, different pedagogies, different learning environments, and so forth, mean that causal inferences about the conditions of design, pedagogy, and technology use are nearly impossible to make with any certainty (p. 1245).” They specifically noted, in support of Clark, that delivery method was often confounded with instructional design.

The Limitations of Prior Research

What existing research does not do is identify which specific instructional methods and instructional activities are more effective than others. Part of this is because of the design of the meta-analyses themselves: Previous meta-analyses did not specifically code for detailed instructional activities, largely because they were interested in comparing media rather than instructional methods. Thus, when the time came to search for moderators, the coding did not exist for detailed analysis of instructional activities.

A second, perhaps more compelling reason for the lack of research on the efficacy of particular instructional methods and activities is articulated by Bernard et al. (2009), who suggested that it would be impossible to draw causal conclusions about the impact of media, pedagogy and other learning environment factors because of the “fundamental confounds” (p. 1245) referred to earlier. The two most glaring of these fundamental confounds are also the most debated in their ontological and instructional effects: the impact of physical presence (i.e., a proximity effect)—or lack thereof—on instruction and the potential augmentation of or limitation of instruction provided by digital computing technologies (i.e., a media effect). Since it is currently impossible to remove either of those as yet poorly understood effects from the distance education context using existing research methods, it is also impossible to separate those effects from studies of the efficacy or efficiency of DE. Moreover, there is a fundamental confound that exists in any instructional context, whether DE or not, between the effects of individual characteristics of the learner and of the instructor on the instructional process itself.

While there may, in fact, be “fundamental confounds” that cannot be separated from each other within any learning environment, meta-analyses structured specifically to compare media are not necessarily structured to detect differences in other aspects of the instructional environment. Given the relative lack of meta-analyses involving instructional technology that specifically investigate the instructional activities or methods used in DE instruction; it may be premature to suggest that those activities and methods are inseparable statistically from other aspects of the instructional environment.

The limitations of prior meta-analyses pointed-out above will be addressed in the present study in three ways:

1) by building on the prior work reported to specifically look for—and code—detailed instructional activities;
2) by comparing the activities rather than the media; and
3) by controlling for as many confounds as possible through the structure of the meta-analysis itself.

This Study: Narrowing the Focus and Scope of Moderating Factors

Previously Identified Moderators

To summarize, the moderator factors that affect student outcomes as identified by the recent large-scale meta-analyses reviewed:

1) Methodological quality and pedagogy (Bernard et al., 2004),
2) Instructor involvement, media involvement and type of interaction (Zhao et al., 2005),
3) Instructional methods (Sitzmann et al., 2009),
4) Time on task and equivalence of instruction (same or different) (Means et al., 2009), and Interaction treatments (Bernard et al., 2009).

Each of these moderators reflect the categories that were coded by a particular meta-analysis and represent groupings of related coded factors that have conceptual similarities across multiple analyses Because these studies were all—with the exception of Bernard et al. (2009)—comparing media delivery as the central comparison for main effect, they were not coded in the detail required to compare actual instructional aspects of the constituent studies. Bernard et al. (2009) approached their study as a comparison not of media, but of specific instructional aspects used within DE instruction. They encountered two circumstances that required them to create large conceptual groups for coding rather than coding for specific instructional activities themselves: a great diversity of described instructional activities and a lack of statistical data for those individual activities that precluded separating one activity from another. This lack of separation creates—whether it actually exists or not—a confounding situation where the effect of two or more activities cannot be separated from each other. As a result, Bernard et al. (2009) grouped instructional aspects of DE into three large categories of interaction that they adapted from Moore (1998): student-student interaction, student-teacher interaction and student-content interaction.

Types of Moderators

Examining the moderators detected by the five meta-analyses listed above, three common moderators of DE instruction were identified: time on task (identified by Means et al., 2009), differences in instructional activity (i.e., “equivalence of instruction” in Means et al., 2009) and methodological quality (Bernard et al., 2004).

Time on task is a moderator of student outcomes that has a long and well-documented research background. The original concept of “time on task” as articulated by Carroll (1963) has essentially been replaced with the concept of “academic learning time (ALT),” which is defined as the amount of time students are successfully covering content that will be tested (Squires, Huitt & Segars, 1983). ALT has been even further identified as referring only to the time during which a student's readiness to learn coincides with an instructional activity that results in actual learning (Aronson, Zimmerman & Carlos, 1999). What is important about this is that “time on task” cannot simply be measured in terms of time spent in treatment as used by Means et al. (2009), but is dependent upon that time being well-used instructionally—that is, time intentionally structured by the instructor to produce student activity that leads to learning (Byrd 2001; Coeyman, 2002). Finally, the effectiveness of time on task is dependent upon teacher competency and requires that learning activities be effectively designed and implemented (Brophy, 1988).

Thus, the time on task moderator may be thought of as a moderator reflecting instructional planning and instructor activity as much as it is a moderator of time spent in treatment. In the case of Means at al. (2009), the time on task moderator was primarily associated with blended classes where the on-line portion of the blended class supplemented rather than supplanted portions of the f2f instruction—this moderator either disappears or is largely unmeasured in studies of on-line instruction. The present study is limited to studies of Web-based distance learning; time on task in such cases is essentially a function of provision for interaction and is dependent as much on the actions of the instructor as on the time spent in instructional activity by the student. It is not separately coded or examined in this study.

Some previous attempts have been made to examine in more detail the difference in instructional activity that might affect student outcomes: Differences in instructional moderators were noted by all five meta-analyses though under different names and studying slightly different aspects of instruction as their focus:

1) pedagogy (Bernard et al., 2004) and instructional methods (Sitzmann et al., 2009),
2) instructor involvement (Zhao et al. 2005) and media involvement (Zhao et al., 2005), and
3) type of interaction (Zhao et al., 2005) and interaction treatments (Bernard et al., 2009).

The moderator indentified by Zhao et al. (2005) as “media involvement” was a coding category indicating whether the study was f2f, blended or DE only; that is, no involvement, some involvement or complete media involvement in the delivery of the instruction. An important element in blended classes is that the media extends and supplements f2f instruction. Thus, as noted earlier, it is fundamentally confounded with time on task. Zhao et al. (2005) used the term “type of interaction” as a coding category to reflect whether student-teacher interactions were synchronous, asynchronous or non-interactive. Bernard et al. (2009) used the term “interaction treatments” to refer to “the conditions or environments that are designed and arranged by teachers to encourage [interaction] behaviors (p. 2010).” It is clear that in both cases the type of interaction identified as being a moderator variable of DE effectiveness referred to actions by the instructor. With the exception of media involvement (Zhao et al., 2005), which is confounded with time on task, ALL the instructional modifiers identified as affecting student outcomes (at a group level) are due to instructor actions. It is important to note that this effect holds for groups, not necessarily for individual students for whom individual characteristics play a substantial role in differences in academic performance. Extending the work of recent meta-analyses on the effectiveness of DE, thus, appears to require drilling down into the specific instructor interventions used in DE instruction. In fact, Bernard et al. (2009) put it this way in the final paragraph of their study:

> If there is any further traction to be gained by conducting DE versus [classroom instruction] CI studies, it is through more refined investigations of how specific instructional methodologies that have proven effective in CI environments such as cooperative learning (Johnson, Johnson & Stanne, 2000) can be adapted for DE. As well, classroom instructors may gain equally from understanding how proven DE practices can successfully be adapted for their use. (p. 1267)

The current study is a first step toward accomplishing just that.

Unlike time on task, methodological quality was controlled for in the Bernard et al. (2009) and Means et al. (2009) analyses, based on its earlier identification as a moderator variable by Bernard et al. (2004). The present study controls for the moderating influence of methodological quality by implementing even more rigorous inclusion criteria than did either Means et al. (2009) or Bernard et al. (2009). Only high quality studies have been included in the meta-analytic sample used in this study.

Identifying Coding Categories

Criteria for Selection of Coding Models

Drilling-down into each of the three moderators requires some conceptual model that subsumes individual activities, but provides greater detail than that afforded by the term interaction. There are three possible options for addressing this need: use existing models or paradigms that are suitable, modify existing models to suit or create and test suitable new models. Of the three options, the use of viable existing models was preferable. Accordingly, a search of the literature was conducted that revealed some likely models and of those models, three were chosen using the following principles:

1) Tested and published models were preferable to untested, unpublished models.
2) Models tested with DE were preferable to models untested with DE.
3) The models had to have elements that were operational in nature and sufficiently described and detailed to act as guides for coding.

Thus, the best models of student or instructor instructional activity within the DE context would be existing models that had been previously published and tested within DE contexts and of sufficient detail to be used as a coding guide with little or no modification. The caveat to this, however, is that the level of coding detail cannot exceed that which the data can support. One drawback to meta-analysis is its reliance upon extant studies: it is impossible to analyze that which does not exist. In the present case, a number of possible models for coding were examined; some were subjected to pilot coding and rejected because the data in the included studies was insufficient to support those models. The final coding scheme utilized relatively coarse-grained categories simply because the available data did not support finer-grained models for coding. This was a problem Bernard et al. (2009), as well as others identified, but the present study still managed to drill-down into the data in greater detail than previous meta-analyses have.

Overview of the Coding Categories

Coding for this study is adapted from models described by Lepp (2010), Maddrell (2008) and Nickel (2010). This adapted coding scheme is used to address instructional interventions pertaining to research question 1. It uses two major categories, Instructional Strategy (IS), which categorizes each study according to the dominant instructional approach used as a contrast in each study and Collaborative Design (CD). Each study was coded according to the type of collaboration designed into the instruction at the center of the study. Studies were grouped according to year in coherent chronological groups in order to answer the second research question.

Design of the Study

The present study "drills-down" into the data provided by Means et al. (2009) and Bernard et al. (2009)—as well as extends it to studies published after their inclusion dates (i.e., July 2007)—by using statistical meta-analysis to study the effects of specific instructional activities (discussed below) on learning outcomes.

This study applies the technique of comparative meta-analysis to identify the most effective instructional activities used for DE as found in the sample of studies in Means et al. (2009), Bernard et al. (2009) and studies published after the cut-off date for inclusion in those two studies (i.e., after July 2008) that meet the same criteria. The basic study design follows the procedure used in Means, but extends it in four ways:

1) by including studies and certain criteria for controlling for research methodology from Bernard et al. (2009),
2) by adding newer studies to the sample (i.e. studies completed since July 2008),
3) by controlling for media of delivery, i.e., Web-based only, and
4) by utilizing a methodology previously developed by the author in an earlier meta-analysis. That study examined the relative effectiveness of various instructional techniques when used in conjunction with particular ways of using a computer (Roberts, 2002). For lack of a better or pre-existing term, this procedure is herein referred to as a comparative meta-analysis.

In brief, the procedure for a comparative meta-analysis is as follows: First, criteria for inclusion of studies is developed based on theoretical grounds and statistical requirements. Second, all available sources for extant studies that meet the criteria are searched. Studies that meet the criteria comprise the study sample. Next, the studies comprising the sample are placed in sub-groupings based on the independent variable(s). The data from each study are subjected to statistical analysis to derive an estimator of effect size *g*, the standardized mean difference (Hedges, 1981). In addition, homogeneity for each subgroup is tested to measure the impact of influences other than the independent variable on the effect sizes. Finally, the fail-safe N (Orwin, 1983) is calculated to determine the adequacy of each sample subgroup.

The results of the initial statistical analysis are ranked according to the magnitude of the effect size and compared to the overall main effect size. Post-hoc factor analyses and other appropriate tests to identify any mediating or moderating variables are conducted. If necessary, new categorical grouping based on the post hoc tests are created and effect sizes, homogeneity and fail-safe *n* are calculated for each newly formed group. Finally, a Binomial Effect Size Display (BESD) (Rosenthal & Rubin, 1982) is calculated for each sub-group to assist in interpreting the effect sizes.

Overall, comparative meta-analytic methodology is used to reexamine and extend the body of research following a design similar to that used by Bernard et al. to:

1) Extend the body of DE studies used in Means et al. and Bernard et al. (2009) to studies published after July 2008 which meet the criteria for inclusion in Bernard et al. A main effect is calculated, using this new sample of studies, in order to provide an overall effect of web-base instruction on student achievement.
2) Compare the effect size of earlier with those completed after July 2008.

3) Drill-down into the moderating factors that Bernard et al. and others identified as having an effect on student outcomes by using finer-grained, theory-based categories than those used in Bernard et al. to compare the effects of various learning activities on student achievement.

This research design is based on three assumptions: First, that Means et al. (2009) and Bernard et al. (2009)—as the most recent and most extensive meta-analyses of distance education studies thus far performed—were sufficiently rigorous that they subsume all previous similar studies; second, that both studies were comprehensive in locating all studies through July 2008 that met their inclusion criteria and that further search for studies prior to August 1, 2008 would be redundant and likely to result in few, if any, additional studies, and third, that their conclusions were sufficiently sound to act as a theoretical starting point for searching for effective practices.

Unlike these earlier studies, however, the current study is limited in several important ways:

1) Only studies involving post-secondary students and adult learners are included; studies involving K-12 students have been excluded, and
2) Only studies involving Web-based distance education are included; studies involving blended (combinations of f2f and DE) instruction and media other than Web delivery have been excluded.
3) Only studies of the highest rigor are included; that is, only experiments and high quality quasi-experiments are included.

4) Only studies in which all necessary and relevant details necessary to conduct a comparative statistical meta-analysis were included in the article; no effort was made to contact authors or publishers to gain statistical data or clarifying information

These important differences mean that the current study is even more exclusive than either Means et al. (2009) or Bernard et al. (2009) and that some studies included in their meta-analyses are not included in this meta-analysis.

Relevance of the Study

The implications for education, and distance education in particular, are important and obvious: First, if instructional method plays a major role in student learning, then educational research should be directed not at trying to differentiate learning according to the media used, but in determining the best instructional practices that lead to the greatest student learning. If learning is contextual as suggested by many (e.g., Means & Haertel, 2004) then distance education may most appropriately be viewed as a specific type of context within which learning and teaching takes place and understanding the best practices within that context may be best considered by examining those instructional practices which, under authentic distance education contexts, seem to lead to higher student learning.

Second, what the instructor does, as well as what the student does, impacts student outcomes. In addition to their mastery of content knowledge, if higher education instructors affect student outcomes through their instructional planning and instructional delivery as seems indicated by the results of previous meta-analyses, then mastery of instructional planning and delivery appropriate to Web-based DE seems to be necessary to student success. This suggests that the implementation of DE services by institutions of higher education requires the participation of faculty trained in course delivery via DE as well as the provision for such training. By extension, this also suggests that higher education faculty should be trained in instructional planning and delivery for f2f classes as well; mastery of content knowledge does not appear to be sufficient in and of itself to assure the best student outcomes: Knowing what works is a prerequisite for such training.

The Research Questions

The two research questions with which this study is concerned are repeated here for clarification:

Research question 1: Which instructional methods are more effective when used in conjunction with Web-based distance education--and under what circumstances?

Research question 2: Have Web-based distance education outcomes improved over time?

Organization of the Study Report

In general, this report follows the Meta-analysis Reporting Standards (MARS) established in the *Publication Manual of the American Psychological Association, Sixth Edition* (APA, 2010). The MARS standards were designed for use in academic journals following American Psychological Association (APA) conventions and do not exactly lend themselves to dissertations, nor to comparative meta-analyses. All of the required content relevant to statistical meta-analyses called for by MARS is included in this report and the headings presented beginning on page 251 of the APA Sixth Edition Manual, as well as the basic sequence of those headings, are followed. Some modifications and additions are made to the content of some sections and some suggested material is not included where it is not applicable to the present study. Presented here is a brief overview of the organization of this report of this meta-analytic study.

Chapter 1 presents the material called for in the section of MARS designated as "Introduction." It includes a statement of the relation under investigation, along with brief versions of the historical and theoretical background leading to the study. It also briefly introduces the selection and coding rationales, as well as the basic methodology used in the study. It concludes with the organization of the study report.

Chapter 2 of this study report, the Literature Review, presents the theoretical arguments leading to the focus on instructional activity as moderator to be investigated—a subject normally covered as part of the methods section of MARS (i.e., Moderator and Mediator Analysis), comparative meta-analysis as the method for pursuing that investigation (not normally included in a meta-analysis report) and the choice of coding categories to organize that investigation. This chapter departs from MARS in that dissertations are somewhat more lengthy and detailed in their theoretical and explanatory aspects than are journal articles. Chapter 2 also includes portions of the MARS methodology section, specifically the preliminary definition of the coding categories. Chapter 2 concludes with a presentation of the coding instrument and the research questions used in this investigation.

Chapter 3 of this study report presents the portion of MARS described under "Method." It describes the comparative meta-analysis process with particular focus on the process used to locate and identify the individual research studies included in the meta-analysis sample (i.e., Inclusion and Exclusion Criteria). Chapter 3 also includes the MARS Method section information listed under Search Strategies, Coding Procedures and Statistical Methods (APA, 2010, p. 251-252). The chapter concludes with the list of included studies and their characteristics and a report of the search and inclusion statistics which MARS lists as part of the results.

Chapter 4 presents the results of the statistical analysis and the findings of the study based on those results. This chapter includes descriptive information on each included study, coding results, grouping descriptives, further modifier investigation and analysis and assessments of bias. The chapter also includes tables and charts illustrating the results of various analyses and a list of the included studies with their associated statistical information. It concludes with a summary statement of the findings of this study.

Chapter 5 includes the discussion of the findings, post hoc explanatory statistical analysis of those findings and conclusions based on them, followed by the implications of those conclusions. It includes most of the information listed under the heading of "Discussion" in MARS. The chapter also includes some non-statistical observations regarding the material and processes encountered in the course of the research for the study and ends with some suggested guidelines for future research.

CHAPTER 2

REVIEW OF THE LITERATURE

This literature review is divided into two parts. The first part contains general `theoretical discussion on the distance education environment, including the roles of technology, the teacher, the student, instructional strategies and assessment. It focuses on the theoretical foundations for what comprises an effective Distance Education Environment. The second part of this chapter is devoted to the recent history of research into the conduct and efficacy of distance education and leads to the theoretical questions with which this study is concerned.

Theoretical Background of Distance Learning Environments

Distance education using modern networked digital computing technologies has become prevalent in the United States over the past fifteen years (Parsad & Lewis, 2008). According to the U.S. Department of Education's National Center for Educational Statistics (NCES), 66 percent of all U.S. 2- and 4-year degree-granting postsecondary institutions offered some sort of distance education instruction in the 2006-2007 academic year (Parsad & Lewis, 2008). In 1995, only 33 percent of those institutions offered distance education courses (Greene et al., 1999). This growth has been accompanied by a commensurate increase in research investigating the efficacy of this type of instruction. These studies on the efficacy of Distance Education (DE) appear to have been predicated on the assumption that there is something qualitatively different about DE in comparison with traditional or face-to-face (f2f) classroom instruction.

Definition of a Distance Learning Environment

The U.S. Distance Learning Association (USDLA) defines distance learning as "The acquisition of knowledge and skills through mediated information and instruction, encompassing all technologies and other forms of learning at a distance (USDLA, 2007)." Thus, distance learning relies heavily, if not totally, on technology as a mediator between the learner, his/her peers and the instructor and is distinguished foremost by its distributed nature, that is, the student and the instructor are never collocated, i.e., in the same location at the same time.

A distance learning environment (DLE) encompasses all the elements of distance learning (which is assumed to subsume such other non-traditional learning systems such as eLearning, Web-based learning, online education, tele-learning and so forth) and is construed as a particular type of a distributed learning environment (American Council on Education, 2001). That includes both World Wide Web-based (Web) and non-Web Internet services and functions and all current and near future methods of accessing the Internet.

The Problems of Researching a Distance Learning Environment

DLEs present a particularly difficult topic of research. Anytime human beings are the subjects of study, high levels of complexity can be expected because human behavior is almost too complex to capture (Kaestle, 1993). Studies involving humans and educational or instructional technology complicate the matter: In addition to dealing with the complexity of human behavior, educational technologists study a field that changes so rapidly that the latest studies are out-of-date before they are published, what Roblyer (2007) calls the “educational technology knowledge gap (p. 1).” The difficulty of addressing a complex, ever-changing environment interacting with a rapid technological trajectory has led to "fragmented and uncoordinated approaches to studying technology resources and strategies (Roblyer, 2007, p. 1)."

In his tongue-in-cheek look at the history of education, Harold Benjamin, writing as "J. Abner Peddiwell" (1939/2004) in *The Saber-Tooth Curriculum,* addressed the problem of educational research, by noting that educational professionals

> . . . required all members of their group to engage in scientific research in education by counting and measuring quantitatively everything related to education which could be counted and measured. . . . [P]rofessors of education . . . confronted almost insuperable obstacles in the fact that education dealt with the changing of human minds, a most complex phenomenon. The task of measuring a learning situation involving an unknown number of factors continually modifying each other at unknown rates of speed and with unknown effects was a tremendous one, but the professors did not hesitate to attack it. (p. 55)

What Peddiwell/Benjamin described in 1939 is what is called a learning environment today. According to Jonassen and Land (2000), learning environments include not only the teacher, the content and the transmissive process, but the learner's activities, the sociocultural and sociohistorical setting in which they act and the tools and mediation systems they use. Reminiscent of Schwab's (1983) four commonplaces of instructional planning (i.e., teacher, student, what is taught and the milieu of teaching-learning) this more holistic view of the learning process has engendered a move away from a focus on the technology used to deliver distance education to a focus on the actions and the context in which distance learning takes place. A considerable amount of dialogue has taken place concerning what constitutes an effective DLE and how the commonplaces—or components—of DLEs function within that environment. Each of those commonplaces or elements of a DLE has a rich background in the literature and are briefly profiled below.

Characteristics of Effective Distance Learning Environments

Defining effectiveness. Defining success or effectiveness in distance education depends upon the operational definition of learning, the purpose of the course and the reason for taking the course. Success also has a lot to do with the perspective of a particular stakeholder. Typically, DE success is seen in differing ways by differing stakeholders: as increased student achievement by faculty and society, as course satisfaction by students, as reduced attrition by program chairs and Deans or as Return On Investment (ROI) by administrators, business leaders and politicians (Gross & Godwin, 2005). Each of these views of success is valid in its own right, but not all are valid in every given situation. Harkening back to the definition of DLE provided by the USDLA (see page 2), learning is the raison d'être of DLEs and thus, the focus of the current discussion.

In general, the measures used historically to determine the success of online instruction have been comparisons to f2f classes using student satisfaction (survey), student achievement (course grades or content related tests), attrition rate, or instructor evaluations (by the student). More effective DLEs have typically been those DLE that compare favorably to f2f LEs on one or more of these measures of success. More effective components of DLEs are those components that compare favorably to either the same component in a f2f environment or another component within the DLE. The findings of numerous studies on the effectiveness of each component of a DLE are summarized below.

More effective student characteristics. Althaus (1997), in a study of 142 undergraduate online students, found that students who were actively involved in computer-mediated discussions earned higher grades than less active students. Important student characteristics that play a positive role in student involvement include motivation and maturity level, prior online experience, college experience or experience in some technical field—but none of these characteristics made a significant difference in the achievement (Benson at al., 2005; Cooper, 2001; Eppler & Ironsmith, 2004; Figuegoa, 1992; Frith & Kee, 2003; Poirier & Feldman, 2004; Thirunarayanan & Perez-Prado, 2001).

Figueroa (1992) compared early online DE and f2f courses in literature in Mexico and found that online students were more personally interested in their learning than were f2f students, who tended to view the course as a school requirement. As a result, the DE students were more engaged with the content of the course than were the f2f students.

Pintrich (2004) developed a theory of self-regulation in learning that posits that some students motivate themselves and need little or no external motivation to be successful learners. Other students are less able to motivate themselves and require external motivation in order to succeed. Highly self-motivated learners are considered to be “self-regulated” and thus tend to be independent learners. The difference between highly self-regulated learners and poorly self-regulated learners may be related to intrinsic versus extrinsic reward orientations and the ability to defer gratification. In effect, self-regulated learners teach themselves and have little need of a teacher while the poorest self-regulated learners rely upon the teacher for even the smallest learning.

Entwistle (2001; Entwistle & Tait, 1990) theorized that all students approach learning in one of three ways at any given time: Deep Learning (learning is primary), Strategic Learning (grade chasers) and Surface Learning (get by). These approaches affect how students perceive effective teaching and are extremely context-based. A student who approaches one learning situation deeply may, in another learning situation, use a surface learning approach. Jelf and Colburn (2002) applied the concept of learning approaches to the use of virtual seminars in a third-year psychology course and identified all three types of learners. They determined that Deep Learners were autonomous, preferred to work independently and were more satisfied with the online virtual seminar environment than were strategic or surface learners. The most salient observation was that, because virtual seminar attendance was not required, Surface Learners chose not to attend because they didn't have to, Strategic Learners found f2f more efficient than DE and Deep Learners liked the autonomy, self-paced atmosphere and the rich material available in online courses attractive. Jelf and Colburn (2002) found no significant difference in the overall perception of online learning or learning with computers in general between learners with different approaches to learning. Thus, motivation and approach to learning can effect achievement regardless of whether learning takes via DE or f2f (Case, Gunstone & Lewis, 2000; Entwistle, 2001; Zimmerman, 1989).

Wolters (1998) found that self-regulated learners are active learners who efficiently manage their own learning experiences. They tend to have large, varied background experience on how to obtain new concepts and apply previous ones to new academic tasks. They are goal setters, intrinsically motivated and willing to be active participants. Self-regulated learners monitor their own progress and are able to make the necessary adjustments that lead to success. They have a high self-efficacy in their ability to succeed and are equivalent to Entwistle's (2001) deep learners. Table 1 illustrates the combined the elements of Pintrich's self-regulated learning, Entwistle's approaches to learning and Wolters' learner characteristics to derive a basic classification for students' classroom goal orientation.

Table 1

Classification of Student Classroom Goal Orientations

Self-Regulation (Pintrich, 2004)	Learning approaches (Entwistle, 2001)	Learner Characteristics (Wolters, 1998)	Student Goal Orientation
High	Deep Learning	defer gratification, persistence, goal driven, self-efficacious, self-confident; learning is valued	Real learning
Medium	Strategic Learning	Short-term goals; extrinsic goals only—class is a means to an end;	Good grades
Low	Surface Learning	Passing grade with least pain and minimal effort	Get by

The student goal classifications are useful in identifying the basic underlying goals that students have when entering a course—whether DE or f2f. These goals inform their behavior, class performance and learning and may have a greater impact on measures of student satisfaction, achievement or instructor evaluation than any other factor in a LE.

More effective instructor characteristics. Blignaut and Trollip (2003) cite the importance of instructor presence in their study of an online course and hypothesized that teacher presence is determined by communicative action in an online environment. Zhao, et al. (2005) found that high instructor involvement and use of both synchronous and asynchronous interactions produced advantages for DE over f2f: “the degree of instructor involvement is a significant distinguishing quality of effective and ineffective distance education programs (p. 1863).” McIssac, Blocher, Mahes and Vrasidas (1999) found that prompt instructor feedback, participation in interactions, encouragement of social interaction and employment of collaborative learning strategies were important to students’ positive experiences in DE courses. Greene and Land (2000) found that guiding questions help students focus their projects, real-time dialogue and feedback with instructors was instrumental in the developing them and student-student interaction, particularly the sharing of personal experience, helped foster conceptual change.

Moderation in group discussions fostered the formation of a community atmosphere in an online course (Winograd, 2000). Knupfer, Gram and Larsen (1997) emphasized the importance of establishing a learning community in online courses. They found that the early establishment of study groups, accompanied by teacher modeling and reinforcing of effective communication, along with the identification of and solution planning for problems all contributed to the success of an online course. Increased interaction resulted in increased learning—test performance, grades and student satisfaction (Bocchi, Eastman, & Swift, 2004). Beer, Slack and Armitt (2005) and Robles and Braathen (2002) recommended that instructors become proficient in group work dynamics.

Other findings from recent studies include: students view e-mail exchanges with instructors as the most valuable learning activity (Frey, Faul & Yankelov, 2003); high interaction and participation are critical to online instruction (Keefe, 2003; Young, 2004); the instructional design of a course is more important than the delivery system in affecting the quality of online discussions and the subsequent learning (Berge, 1999).

Online pedagogy seems to come naturally to some instructors but not to others according to Hansen and Gladfelter (1996). They suggested that focus on lectures and text readings while neglecting the creation of respect and safety was detrimental to productive debate and collaborative problem-solving. The delivery of instructor-based training should be responsive to individual student learning differences (Boyle, Kolosh, L'Allier & Lambrecht, 2003).

Online instruction can produce academic achievement superior to f2f under certain conditions, particularly for traditional lecture courses (Maki, Maki, Patterson & Whittaker, 2000). In such cases, students in online courses typically have access to additional materials and the classes are extended beyond the meeting times through on-line student-student contact and opportunities for student-teacher contact not available to f2f students. Moreover, taped lectures can be viewed repeatedly to glean information missed the first time around; f2f students have one shot at it. This suggests that on-line versions of course conducted using the traditional lecture/reading/writing/mid-term/final instructional practices are superior to the f2f versions. The on-line environment requires taped lectures and written communications between teacher-student and student-student in the case of course management systems where student e-mail contact through the CMS is automatically provided.

Lee Shulman (1986), in his Presidential address at the 1985 Annual Meeting of the AERA, points out that the etymological roots of our highest academic degrees "master" and "doctor" both derive from the concept of "teacher" (p. 6) and the universities that grant them are descended from normal schools whose task was preparing the highest level of scholar: the teacher. Shulman bemoans the fact that somewhere during the long history of teachers, teaching became divided into content knowledge and pedagogical knowledge, though originally no such division existed (see Ong, 1958).

Shulman (1986) lists and describes the three types of knowledge that a teacher should possess:

- *Content Knowledge* (Domain knowledge—what a teacher should know)
- *Pedagogical Content Knowledge* (Instructional skill—what a teacher should do)
- *Curricular Knowledge* (includes instructional technology (IT), by definition, but perhaps IT belongs more in Pedagogical content knowledge—what a teacher teaches with, a teacher's tools)

Similarly, Zhao at al. (2005, p. 1861) found three interaction-related factors related to effective distance education: instructor involvement, particularly in the actual delivery of content, media involvement and types of interactions.

The three types of knowledge that teachers must possess closely match the interactions that relate to effective DE. These two lists are combined in Table 2 to derive a third list that illustrates how the three types of teacher knowledge address the three interaction-related factors for effective DE.

Table 2

Teacher Knowledge, Instructor Involvement and DE Teacher Competencies

Teacher Knowledge Types	Instructor Involvement	DE Teacher Competencies
Content Knowledge	Actual delivery of content	Mastery of Domain knowledge
Pedagogical Content Knowledge	Managing interactions	Plans, moderates and participates in meaningful interactions
Curricular Knowledge	Media involvement	Deploys technology appropriately

In other words, it seems online instructors must be masters not only of content knowledge but of pedagogical content knowledge of best practices, particularly those best for online pedagogy and curricular content knowledge, including technological content knowledge.

More effective instructional strategies. Interaction is “the single most important activity in a well-designed distance education experience,” according to McIssac, Blocher, Mahes and Vrasidas (1999, p. 122). Note starters appear to help students see differing points of view during online discussions and helped move them away from mere statements of agreement to actual discourse that included disagreeing with other students statements (Nussbaum, Hartley, Sinatra, Reynolds & Bendixen, 2004). The use of *emoticons* has been successfully used to compensate for missing visual and nonverbal communications cues in online communications (Bielman, Putney & Strudler, 2000). Gilbert and Dabbagh (2005) found that facilitation and evaluation guidelines for discussion postings resulted in deeper and more meaningful student content learning.

More effective assessment. Gayton and McEwen (2007) determined that effective techniques for assessing online student learning have not been thoroughly addressed while Robles and Braathen (2002) concluded that online instruction requires a more on-going, systematic approach to assessment than that used with traditional instruction. Liang and Creasy (2004) discovered that using online assessments caused instructors to modify their methods of instruction, often having to become more innovative than would otherwise be the case in traditional instruction. Multiple assessments should be utilized in an online environment (Christopher, Thomas & Tallent-Runnel, 2004; Gayton & McEwen, 2007; Robles & Braathen, 2002).

More effective uses of technology. How important is technology in the learning environment? Lant (2002) believes that online technologies and face-to-face instruction are complementary. This suggests that blended classes have the potential for leveraging the best aspects of both DE and f2f instruction to create the most effective LE. Levin, Levin and Waddoups (1999) suggest that online instructional formats may require the creation of new ways of learning and new methods of teaching, including innovative evaluation methods.

Levin, Levin and Chandler (2001) suggest that some digital communicative tools such as video conferencing can effectively “remove” distance and create effective social organizations Zhao et al. (2005, p. 1863) noted “the use of technology to remove the distance between the provider and recipient of instruction." Notice the clear influence of the transmissive view of instruction in the latter study and the technological determinism present in the first study.

Jackson and Wolski (2001), in a study of students' pre-instructional beliefs about science, found that interactive online technology provides a reliable method for controlling and preserving student dialogue in order to extract and analyze student's points of view. Im and Lee (2003/2004) found that synchronous communications promoted social interaction best, while asynchronous was better for task-oriented communication.

In an early study of online courses, Christel (1994) found that video presentation of important pedagogical information led to better recall than other presentation methods. A caveat to Christel's findings was noted by Mayer, Heiser and Lonn (2001) who found that animation accompanied by narration created a high cognitive load that inhibited transfer of complex concepts. In a later study, Mayer and Chandler (2001) found that a modicum of interactivity accompanying multimedia presentations helped overcome cognitive load and fostered deep learning.

Bee and Usip (1998) showed that the online provision of supplementary materials (whether in online or f2f formats) improved student performance, but only when students used them. Ahern and Durrington (1995) found that anonymous communications fostered highly structured communication patterns (longer messages and more time expended on constructing them) and, when combined with graphical interfaces, encouraged students to engage in highly structured interpersonal interactions. Asynchronous communications can promote learning (Bodzin & Park, 2000; Jonassen, Davidson, Collins, Campbell &

Bannan-Haag, 1995; Kaye, 1992; Vonderwell, 2003). Meaningful discourse is constructivist and social and promotes learning through articulation, reflection

The benefits of online technologies include electronic grade books that give faster student access to results, allow for measuring learning more accurately, and helps foster a student-centered learning environment (Bartlett, Reynolds & Alexander, 2000; Farmer, 2005; Liang & Creasy, 2004).

Summary. Robles and Braathen (2002) found that online education alters the way humans interact, causing them to modify their methods of communication, learning, and assessment. In their meta-analytic study, Tallent-Runnels, Thomas, Lan, Cooper, Ahern, Shaw and Liu (2006) concluded that there was sufficient evidence from multiple studies to support the following generalizations about effective online course instruction and management practices. Effective courses include the following characteristics:

- the creation of learning communities through the formation of small groups
- the modeling of effective communication by the instructor
- instructor presence created through active participation in discussions, timely feedback, and frequent announcements
- instructor scaffolding of discussions
- the promotion of and participation in teacher-student and student-student interaction
- teacher interaction that reflects deep understanding of course content

In conclusion, the following points seem to differentiate more effective distance (online) learning environments from less effective distance education practices:

- Web-based courses require more teacher-student contact than f2f.
- Web-based courses require intentional group-building and social networking skills on the part of instructor.
- Online learning appears to require more individualized instruction than does f2f.
- Interaction in Web-based courses should include timely, frequent and meaningful personalized feedback.
- Web-based courses seem to benefit from guided questioning, active participation, instructor modeling of good discussion practices and scaffolding from low quality to higher quality responses in online discussions by the instructor.
- The technology used to deliver Web-based instruction matters less than developing relationships, meaningful interaction and creation of a safe and non-threatening community.
- Instructor presence in Web-based courses must be intentionally created and maintained.
- Both synchronous and asynchronous communication are important to building community and creating instructor presence (this means the instructor and students have to "meet" online at the same time.
- Variety in assessments and assignments was named by both faculty and students as important to learning and the online experience.

- The instructor must actually deliver instruction in an appropriate form and format, rather than merely post materials and assignments on-line.

All these to overcome the proximity effect of f2f; what emerges is a model of DE courses at odds with a very common model where the course readings and assignments are posted on a course management system along with links to appropriate online tutoring or help material and students only contact the instructor when problems occur and never interact with each other. This model is, in effect, online delivery of independent study. The key difference seems not to be the content of the course, nor even the course assignments or requirements, but the presence of the instructor and the development of relationships that lead to content-oriented interaction between teacher and each student and among the student themselves.

Conclusion. In the final analysis, good teaching is good teaching. What appears to most effective in f2f settings also seems to be most effective within DLEs and, conversely, what doesn’t work in one setting doesn’t work in the other. The major difference between DLE and f2f formats, thus, seems to be the role of technology as a reducer of distance and a mediator of communications. The presence of technology adds a requirement for technological skills necessary for both teachers and students to maximize the mediation and distance reducing capabilities of digital computing technologies. This suggests that what is most important in DE research is not trying to determine whether DE is more or less effective than f2f, but identifying the maximally effective ways to teach using digital computing technologies followed by development of ways to impart that knowledge to teachers who teach in DLEs.

The History of the Study of Distance Education

The "No Significant Difference" Phenomenon

Historically, studies comparing distance education to f2f instruction have found that there was no significant difference between distance and f2f instruction. This tendency has become known as the "No Significant Difference" (NSD) phenomenon (Russell, 1999). The NSD phenomenon can be attributed to a number of causes: viewing instruction as transmissive, viewing technology as determinative, failing to recognize the emerging nature of instructional technology implementations, attempting to study a complex environment in a purely reductionist fashion and inadequate research design and implementation. Regarding the latter, Phipps and Merisotis (1999) listed several systematic factors that may account for why NSD or discrepancies in findings occur in studies of DE: small sample sizes, high attrition in online courses, inadequately designed and tested data-collection instruments, nonrandom sampling, failing to consider independent variables such as age, gender, experience and lack of reflexivity by participants or researcher.

The transmissive view of instruction. The historically prevailing view of instruction is transmissive, that is, knowledge is an object (epistemologically-speaking) that can be transmitted from teachers to learners (Kember & Gow, 1994). It is based upon a communications model involving a sender (i.e., the teacher), a receiver (i.e., the learner), a message (knowledge) and a transmission medium (instructional methodology plus instructional technology) (Lasswell, 1971). Under this model, good teaching is good communicating and good communicating involved a "meeting of the minds" (Holmes, 1897) coupled with reduction or elimination of transmission interference (static)

(Jonassen & Land, 2000). The transmissive model of education was made more pedagogically effective by the introduction of a feedback system (again, a communications concept) that provided a check for whether the message was being accurately received or not (Vines & Rowland, 1995). The role of instructional technology within the transmissive model of education was as a medium to carry the message (McLuhan, 1994). Research was focused on reducing the effects of the medium on the message—that is, eliminating or reducing static. It is this model of education that Clark (1984) had in mind when he labeled instructional media a “delivery truck.” To the extent that education is transmissive, the delivery truck metaphor is accurate.

Marshal McLuhan (1994) introduced a conundrum into the transmissive model with his notion that the “medium was the message”—that is, one cannot separate or divorce the message from the medium used to transmit it and any given medium altered the message in some way (p. 9). The reverse, though less quoted, was also true according to McLuhan: the content of a medium blinds us to the character of the medium (McLuhan, 1994). Though he was largely referring to mass communication—specifically television—the concept was readily applied to computers when they began to be used in education and, later, to the Internet as well. For McLuhan, transmissive media was an extension of man’s own transmissive capabilities—visual, auditory, tactile, kinesthetic, and olfactory. Media amplified or extended those communicative abilities. Naturally, instructional technologists became concerned that, if true, McLuhan’s contention that the medium and the message altered each other might mean that instructional technology automatically added—or subtracted—information to the transmission from teacher to learner. So, researchers began investigating whether—and how—instructional technology affected the transmission of instructional messages and, if so, whether it affected student learning (i.e., achievement). Some researchers and theorists, like Jonassen (2000b), considered the possibility that instructional media and technology could augment or improve transmission, much as filters and amplifiers could improve the electronic transmission of a signal (*cf.* Salomon & Perkins, 2005). Recently, however, a paradigmatic shift seems to have taken place in the understanding of the teaching-learning process (Barr & Tagg, 1995). Jonassen (2000a) contrasts the transmissive model with the new paradigm in which learners work on authentic, contextualized (real-world) problems that are ill-structured and ambiguous (with multiple embedded issues and solutions).

The emerging nature of instructional technology implementations. Past research on DLEs has been limited partially because past researchers have studied technology at an emerging stage of development. There is a huge difference between current computer technology and past technologies available for use in earlier implementations of distance education. More than that, most research has necessarily focused on DE as implemented, rather than DE as possible. Zhao et al. (2005) agree: “Either because of the limitation of technology or because of cost, distance education programs, until recently, have not been able to offer the full range of communication channels to students and instructors (p. 1862).”

The best DE programs will leverage both the maximum possible potential of technology and instructional practice as well as the intersection of the two. Thus, the real question is whether it is possible to compare maximally implemented DE instruction with maximally implemented f2f instruction. In other words, have the implemented DE and f2f examined in the past truly reflected the maximum potential of either mode of instruction or merely some lesser implementation? If the latter is the case, then the actual observations made by of past studies have been of the relative weaknesses of the implementations and not of their relative differences.

Inadequate design and implementation. A final reason that so many studies fail to find any differences between DE and f2f is simply that they are inadequately designed and implemented. Roblyer (2005; 2006) has recently drawn attention to the fact that far too many studies of instructional technology and online education are not rigorous, lack evidentiary bases, have weak research designs and are poorly written. In their large literary review of literature on studies of online teaching, Tallent-Runnels, et al. (2006) found that studies of the online environment were descriptive in nature, utilized small, non-random samples and often studied unique groups or specialized programs. They also found that blended or hybrid courses were frequently labeled “online” or “distance education” and most studies of online learning situations lacked empirical data with which evaluations of the effectiveness of assessments or procedures could be checked by peers.

Recent Research Contrasting DE and F2f

Three recent, large-scale meta-analyses of studies comparing face-to-face and distance education support the historically consistent finding that no significant difference in student outcomes exists between distance education and face-to-face courses (Bernard, Abrami, Lou, Borokohovski,Wade, Wozney, Wallet, Fiset & Huang, 2004; Means, Toyama, Murphy, Bakia & Jones, 2009; and Zhao, Lei, Yan, Lai & Tan, 2005). All three studies concluded that *factors other than the media used to deliver instruction affected student outcomes*. A large-scale study of distance education performed for the Canadian government (Ungerleider & Burns, 2003) and a later, smaller meta-analysis performed at the University of British Columbia (Jahng, Krug & Zhang, 2007) also concluded that no significant difference between f2f and DE existed. In other words, the aggregate finding of nearly eight decades of research on the efficacy of distance education has consistently shown that, as asserted by Clark (1983, 1994), media seems to be irrelevant—that is, there is no generic media effect on learning detectable by current methods of research.

Significant Heterogeneity between Studies

Unlike earlier studies however, each of these more recent studies also noted a peculiar phenomenon: While the aggregate effect sizes for distance education compared to face-to-face showed no significant difference in student outcomes, significant differences were found within each group. Specifically, individual studies of distance education learning differed widely from each other and the same was true for the face-to-face studies that the meta-analyses examined (Bernard et al., 2004; Means et al, 2009; and Zhao et al, 2005). This effect was also noted in separate individual studies by Keefe (2003), Poirier and Feldman (2004) and Campbell, et al. (2008).

The finding of no significant difference across groups and significant differences within groups typically indicates that some factor or factors—either mediating factors or moderating factors—other than the treatment are affecting the outcomes. In meta-analyses, when faced with such a statistical condition, the accepted practice is to look for moderator factors (Hedges, & Pigott, 2004; Means et al., 2005; Shadish, & Sweeney, 1991; Zhao et al., 2005). Accordingly, each of the aforementioned meta-analyses conducted statistical searches to identify any factors that may have affected the student outcomes. Bernard and colleagues (2004) examined attitude, retention outcomes and asynchronous/synchronous formats. Zhao and colleagues (2005) looked for differences in publication features (year and author), study features (design, measurement, results, etc.), instructor involvement and status (i.e., Professor, graduate student, etc.), learner features including background and status, content area, class time, credit type, course setting (professional, K-12, graduate, etc.), media involvement and interaction type (asynchronous/synchronous, both or none) between studies of DE. Means and colleagues (2009) looked at twenty-one different coded factors grouped as practice variables (pedagogy/learning experience, synchronous/asynchronous communication, treatment duration, presence of multimedia, time on task, presence of face-to-face opportunities, practice opportunities and feedback), study conditions (year of publication, learner type and subject matter) and study method (sample size, type of knowledge tested, study design, unit of assignment to conditions, instructor equivalence, and equivalence of curriculum/instruction.

Identifying Moderator Variables

Zhao et al. (2005) found that instructor involvement, media involvement and the type of interaction were factors that moderated student outcomes in the studies included in their meta-analysis. Means et al. (2009) found that, of the twenty-one factors tested as moderators, only two factors emerged as statistically significant moderators: time on task and equivalence of curriculum and instruction. In a narrative review of studies that did not meet the criteria for inclusion in their statistical meta-analysis, Means et al. (2005)

also found that only learner control of media interactions and support for meta-cognitive regulation seemed to positively affect student outcomes and that providing learning guidance seemed to be more effective with individual learners than with groups.

The presence of media involvement as moderating factors in both the Zhao et al. and Means et al. (2009) studies seems to contradict the overall finding that media does not make a difference in student outcomes. This is misleading because it is the type of interaction students have with the media, rather than the media they have access to, that seems to make the difference. This is in line with the theories of Jonassen (1996, 2000) who asserts that it is how technology is used that determines its effect on learning rather than simply that technology is used. In a meta-analysis investigating the most effective instructional use of computers, Roberts (2002) found support for Jonassen's theory. More recently, Zhang et al. (2006) in a study of various distance education media concluded that how a medium is used is more important than simply having access to it.

The moderating factors identified by Zhao et al. (2005) and Means et al. (2009) as having the greatest affect on student outcomes in studies comparing distance and face-to-face instruction are among those that have separately been identified as having a beneficial effect on student learning in comparisons of various instructional interventions involving solely face-to-face instruction (Admiraal, Wubbels & Pilot, 1998; Black, 2003; Brophy, 1988; Brown, 1987; Flavell, 1979; Hartman, 2001; Long, 1983; Pintrich, 1988; U.S. Department of Education Office of Educational Research and Improvement, 1994; Zhang, 2001). This suggests that instructional interventions that work best face-to-face also work best in a distance education setting. This makes sense given the assumption—which seems to be valid based upon the evidence from the preponderance of extant studies—that the medium used does not affect student outcomes. Thus, the question of whether or not distance education can be an effective instructional delivery method appears to be the wrong question. Rather, the question that should be addressed is: Which instructional interventions are most effective when used in a distance education setting--and under what circumstances?

Sabelli (2004), in a study of instructional technology research patterns, concluded that researchers tended to overlook the importance of the teacher and instructional method when assessing the impact of technology on student learning. This is in accordance with the findings of Means et al. (2009) that equivalence of curriculum and instruction was a significant moderating factor on the effect size of student outcomes within groups of studies.

Clark (1984) argued that the results of media comparison studies were confounded by differences in the instructional methods used, making it impossible to determine the true cause of differences in student outcomes between mediums. In another recent meta-analysis comparing the effectiveness of web-based and classroom based instruction, Sitzmann, Kraiger, Stewart and Wisher (2009) found that instructional methods were the source of differences in the effectiveness of the studies included in their meta-analysis. In addition, they found that Web-based courses tended to use a greater variety of instructional methods than face-to-face courses and also tended to require students to be more active in their learning (p. 29).

In a 1980 meta-analysis, Kulik, Kulik and Cohen examined the effectiveness of computer-based college teaching and concluded that:

> . . . only one variable predicted study outcome in our meta-analysis, and that was use of a design that controlled for instructor effects. In studies in which different teachers taught computer-based and conventional sections of a course, examination differences were more clear-cut and favored computer-based teaching. In studies in which a single teacher taught both experimental and control classes, differences were less pronounced. . . . It seems possible that involvement of teachers in innovative approaches to instruction may have a general effect on the quality of their teaching. (p. 539)

Kulik, Kulik and Cohen—inadvertently or otherwise—succumb to technological determinism here and attribute the increased quality of teachers' instruction to the effects of using innovative media. It is far more logical to assume that it is the tendency of high quality teachers to investigate innovative media in an effort to improve the quality of their instruction and that they became more effective teachers precisely because of that tendency. In other words, rather than the media creating better teachers through use, better teachers may be more likely to use new media than less effective teachers.

However, even better teachers may not be able to produce better student outcomes when using innovative media. Long and Jennings (2005), in a randomized, controlled study of the effects of an electronic field trip program found that the effectiveness of the program was directly tied to teacher knowledge of, and experience with, the program. This implies—as one would expect—that the effect that media may have is to reduce the effectiveness of teachers during the time required to master the media. This is an area of research that has not received much attention and cannot be further addressed here.

The Focus of Moderator Research

Thus, there seems to be agreement among these separate studies that media—in and of itself—has little discernable effect on student outcomes, rather instructional methodology and other instructor effects have the greatest impact on learning. The role of media, then, seems to be that of merely one more part of the instructional context, a role that is largely dependent upon how it is used as an instructional tool for its effect on outcomes. What the existing research does not do is identify which specific instructional methods and instructor effects are more effective than others.

Identifying Progress in DE Instruction

Given the large number of studies that have investigated the potential effectiveness of distance education, it would be natural to conclude that that research has resulted in some sort of improvement in instructional practices. Quite apart from the possible improvement in the instructional aspects of distance education, is the undeniable improvement in the ability of modern technology to mediate distance and close the gap between teacher and learner. The combination of improved technological capability and improved instructional methods to utilize that improved technology should logically result in some sort of measureable improvement in the outcomes of DE. Very few research studies have specifically studied any sort of trend analysis of DE instructional efficacy. Zhao et al. (2005) compared the effect size for studies in their sample that were published prior to 1998 and those published later. They found that the average mean effect size (Cohen's *d*) for DE studies conducted prior to 1998 were significantly lower than those conducted between1998 and 2001 (-.10 and .20 respectively). The authors speculated that this impressive increase in DE efficacy could be attributed to four factors: more powerful delivery media, more sophisticated support systems, maturation of distance education programs (which includes better trained, more experienced instructors and students who had become more comfortable with online learning) and the technologies used to deliver it. They also acknowledged the possibility of “a paradigm shift” in which distance education had become more “accepted as an effective form of education leading to only studies with positive reports [being] published” (i.e., publication bias) (Zaho et al., 2005, pp. 1864-5).

Early Progress Not Sustained

Table 3 compares the findings of eight meta-analyses published since 2003. At first glance, not much progress seems to have been made since the initial improvement reported by Zhao et al. In 2005, Zhao and colleagues reported an effect size of .20 for studies published in 2001. Bernard et al. (2009), after adjusting for study quality, calculated an effect size of g = .38 and an unadjusted effect size of g = .10, but effect sizes for meta-analyses in between centered-around “no effect” whatsoever. In fact, Ungerleider and Burns (2003) found a zero effect (that is, no difference whatsoever between f2f and DE).

Making Sense of the Progress

Three observations are necessary to make some sense of the findings reported above. First, all of these studies, except Bernard et al. (2009), were comparing f2f to DE—so the effect sizes were not measuring the total effect of DE, but the differences in effect between f2f and DE instruction. Thus, any gains made in instructional efficacy for DE delivery would be offset by any commensurate gains made by f2f instructional practices.

Second, statistical considerations such as differences in effect size measures reported (*g* versus *d*—by its nature, *g* is always more conservative than *d* and thus, somewhat smaller for the same data) or diverse inclusion procedures and widely varying sample sizes. Third, while the foregoing discussion assumes that the no significant difference finding accurately suggests that there is no generic media effect, there is the possibility that media does affect instructional choices and induces users to make choices about their behavior that do affect learning outcomes. Thus, the presence of diverse media delivery systems and methods that are mixed together in both individual studies and in the meta-analyses reported in Table 3 may have some measurable effect on the aggregate effect size reported by each meta-analysis.

Future Directions for Research

Instructional Methods

Some general conclusions regarding the state of DE at the beginning of the second decade of the Twenty-first century may be derived from the meta-analyses mentioned above. First of all, media comparison studies have served their purpose in pointing the way to the next generation of studies. As Collins (2000) suggests, "simply comparing student performance in Web and traditional courses is not the best way of deciding on the success of such new approaches. However, such a comparison should be considered as a *first step* [emphasis added] (p. 26)." The next step, as suggested by Bernard and Abrami (2004b), might be examining various instructional strategies for achieving simple knowledge, comprehension, or higher order thinking skills in online instruction, such as problem-based learning and collaborative learning (p. 416). Likewise, Sitzman et al.

Table 3

Comparison of Recent Meta-analysis Results

Study Name	Years covered	*k*	*d*	*g*	SE	Qτ	Lower limit	Upper limit
Zhao et al. 2005A	before 1998	20	-.10	-	.11	24.400	-.01	.72
Zhao et al. 2005B	1998 - 2001	77	.20	-	.04	484.560	-	-
Lou, Bernard & Abrami 2006	through 2002	218	-	.016		824.348	.012	.044
Bernard et al 2004	through 2002	318	-	.013	.01	1,191.320	-.0068	.0325
Jahng et al 2007	1999 - 2003	20	.023	-	-			
Ungerleider & Burns 2003	2000 - 2003	12	-	.00	-			
Sitzman et al 2005	through 2004	71	.15	-	.02	267.490	.11	.19
Means et al 2009	through mid-2008	28	-	.14		145.58	-.80	1.11
Bernard et al 2009	through 2008	74	-	.10/.38*	.03	209.86	-	-

* unadjusted/adjusted for study quality

(2005) concluded that “instructional methods may be more important than delivery media for ensuring effective learning (p. 29).” There seems to be some level of agreement suggesting that the next stage in studying the effectiveness of DE involves looking at what specifically makes it effective.

The Role of Collaboration

Another direction for research indicated by the results of recent research is the role of collaboration in effective DE. In a meta-analysis following-up on the findings of Bernard et al. (2004b), Lou, Bernard and Abrami (2006) found that “effect sizes are significantly heterogeneous, suggesting that the mean effect size may not be representative of the findings integrated and that other study features may moderate the magnitude of the effect sizes (p. 158).” Specifically, they noted that though DE outperformed f2f when interaction between students and between students and the instructor increased, the individual effect sizes were highly heterogeneous “indicating that although some types of discussions were effective on student achievement, some were not (p. 164).”

Increasing Effectiveness

A third line of inquiry that seems to be suggested by the research to date is whether or not any progress is being made. For instance, as noted earlier, Zhao et al. (2005) noted that the year of publication of studies in their meta-analysis was a significant moderating factor in the outcomes. They found that in studies published before 1998 there was no significant difference between DE and F2f, but in those published in 1998 and later they detected an advantage for DE over f2f and suggested the DE was “getting better” (p. 1055). However, Means et al. (2009) found NSD in effect size between studies published before 2004 and those published between 2004 and 2007. This would seem to imply that whatever improvement that had taken place earlier might have reached a plateau. Both suggested that future research investigate whether or not on-going research was making any difference in the effectiveness of DE.

Research Quality

A final area of concern, but not necessarily for research, was the near unanimous observation regarding the poor quality of research on the effectiveness of DE. Ungerleider and Burns (2003) lamented on how few useful studies they were able to locate (8 out of 11,556) from the years 2000-2003, and observed that "less than a third of the studies devoted to online and networked learning that we identified and reviewed made use of control or comparison groups. We regard this as a significant shortcoming in the research (p. 42)." Bernard et al. (2004a) urged that future research reports "employ more rigorous and complete search methodologies, including more detailed description of control conditions in terms of both pedagogical features and media characteristics (p. 416)." Lou, Bernard and Abrami (2006) recommended that future research provide complete descriptions of instructional conditions, complete descriptive statistical information and descriptions complete methodological procedures, particularly of classroom conditions. These concerns were echoed by others (e.g., Bernard et al., 2009; Jahng et al., 2007; Means et al. 2009).

Statement of the Research Questions

Following up on the above suggestions for research, the current study seeks to provide a preliminary identification of those instructional practices and methods that appear to be more effective when used in a distance education setting. As the recent research syntheses reported above, the current study will employ statistical meta-analysis to analyze both the data already examined by the foregoing research syntheses, but also research that has taken place since the publication of the last of those research syntheses.

In essence, past research on instructional technology has focused on media behavior rather than human behavior. This study will investigate the human behavior associated with instructional media use that affects learning outcomes. It will do so by examining the body of research conducted to this point in time using a different lens than has been used in the past. As part of that investigation, this current study will investigate how different types of collaborative design used in DE affect student achievement. Finally, this study will examine the chronological trends in DE research to determine whether or not any progress has been made on improving the effectiveness of DE outcomes. These research goals are formally articulated as two guiding research questions used in this dissertation:

1. Which instructional methods are more effective when used in conjunction with Web-based distance education?
2. Have Web-based distance education outcomes improved over time?

The next chapter, Chapter 3, describes the methodology of this study, including the criteria for inclusion in the meta-analysis, the creation of coding categories appropriate to answering the two research questions posed here, and the statistical methodology used to analyze the coded studies. Chapter 3 will culminate with a list of included studies and appropriate study-level statistics used to generate the statistical results used to answer the foregoing questions.

CHAPTER 3

METHODOLOGY

Chapter three appears in three parts: It begins with a discussion of the methodology used in this particular study and moves to a description of the meta-analytic method itself. Next, the chapter describes the selection and coding process whereby the criteria for selection and the coding instrument that was introduced in Chapter 2 are used to identify the studies included in the meta-analytic sample and then to extract from that sample the relevant data. The chapter concludes with a presentation of the list of included studies (the sample), their categorization and the individual study statistics used in the meta-analytic analysis.

Restatement of Purpose

The purpose of this dissertation is to explore how educators can more effectively use distance education to meet the educational needs of 21st century students. The preceding literature review suggests that historically, research syntheses on this subject have tended to compare distance education to face-to-face (i.e., traditional) instruction without taking into account numerous modifying and mediating factors that affect the outcomes of distance education. More recently, authors of research syntheses have recognized the need for addressing the impact that various factors other than delivery medium have on the synthesis outcomes. Among the factors affecting the outcomes in distance education research that have not been widely studied using research syntheses, is the impact of differential instructional methodologies on student achievement in distance education. This dissertation addresses that lack.

The assumption made here, based upon the Review of the Literature, is that different instructional methods have a differential effect on student learning, and that those effects may be different when used as part of distance education courses than when used in face-to-face courses. If this is, in fact, true then the resultant effect on the measurement of student achievement in distance education classes will likely vary according to the instructional method used. This variance should be accompanied by a corresponding change in the estimated effect size for each instructional method. By comparing the effect sizes of the instructional activities used in distance education, some idea of which instructional techniques are more effective for student learning can be estimated. In this way, the aforementioned gap in empirical research addressing this particular moderating effect on DE outcomes can begin to be closed. A statistical meta-analysis, as described later in this chapter, was conducted to determine the effect sizes used to make those comparisons.

Research Questions

This study uses statistical meta-analysis as an analytic procedure to estimate the effect that various instructional activities have on student learning when those activities are used as part of Web-based distance education instruction. The magnitude of the estimated effect for each type of learning activity is compared to the others in order to produce a rank-order list of instructional activities according to effect. This was accomplished by extending and drilling-down into the results of several recent comprehensive meta-analyses comparing studies contrasting face-to-face, blended and Web-based distance education instruction (e.g., Means et al., 2009). This two-part investigation, therefore, addresses two questions:

Research question 1: Which instructional interventions are most effective when used in a Web-based distance education setting—and under what circumstances?

Research question 2: Have Web-based distance education outcomes improved over time?

Research Question 1

The first research question addressed by this study is: Which instructional methods are more effective when used in conjunction with Web-based distance education? To answer this question the study population for Means et al. (2009) and Bernard et al. (2009) are used as starting points to identify extant studies that investigate the effect of Web-based instruction on student achievement. Additional searches were made to locate as many similar research studies as possible not already identified in those two meta-analyses. Studies are included in this meta-analysis according to criteria detailed later in this chapter.

All studies meeting this inclusion criteria were coded for two categories of instructional activity (Instructional Strategy and Collaborative Design) used. Groups of like studies were formed for each identified type of Instructional Strategy and Collaborative Design and an effect size for each group completed. These effect sizes were compared to each other and the groups placed in rank order. The group with the largest effect size is considered to be the most effective instructional strategy for use in a Web-based DE environment. No hypothesis regarding what they instructional groups or which is more effective was formulated on purpose. The intention is to let the data “speak for itself” and to keep the researcher as neutral as possible.

Study quality will not be addressed or coded in the current study; it is assumed that Means et al. adequately controlled for that moderator and, by using the same selection criteria, so does the current study. Likewise, time-on-task is controlled for by the selection criteria, effectively eliminating each as a factor in the outcomes measured by this study.

Research Question 2

The second research question this study investigates is: Have Web-based distance education outcomes improved over time? To answer this question, all studies meeting the inclusion criteria were grouped according to their date of publication or, in the case of unpublished studies, the date of the completion of the report of that research, using the three date groups detailed earlier. An aggregate effect size was calculated for each group and the effect sizes compared to each other. It is hypothesized that the aggregate effect size for studies published in Group 3 (2009 – 2010) will be larger than the effect size for the aggregate group of studies in Group 2 (2006 – 2008) and that the effect size for Group 2 will be larger than the effect size for Group 1 (2005 and earlier).

Independent and Dependent Variables

The independent (or predictor) variables of interest in this meta-analysis are Instructional Strategy (IS) and Collaborative Design (CD). The dependent or outcome variable in each case is Student Achievement (SA). Instructional Strategy is defined, for the purposes of this study, to be any activity other than collaboration that takes place within the context of Web-based distance education that fits one or more of the descriptions of activity found in the coding list. Likewise, Collaborative Design is defined as any activity specifically described as collaborative in nature that takes place within the context of Web-based distance education and that fits one or more of the descriptions of collaboration also found in the coding list. Student Achievement is defined to be some quantifiable change in performance that results from a manipulated treatment under experimental or quasi-experimental conditions and that can be statistically analyzed and numerically expressed. In most cases, the exact nature of the achievement being measured within each constituent study is unique to that study but is, in all cases, quantitatively measureable.

Outline of Method: A Statistical Meta-Analysis

Overview of a Meta-Analysis

There is a body of extant research studies describing various distance education courses in which each includes a description of—or identification of—the instructional method(s) used and the outcomes of the course. Researchers have long employed secondary analyses called, generically, research syntheses, to evaluate the cumulative interpretation of similar bodies of research. Detailed discussion of research syntheses appears in Chapter 2 and will not be repeated here but, briefly, a research synthesis strives to "allow the researcher to see patterns across studies that are not apparent when studies are examined individually or serially" (Cooper & Hedges, 1994, p. 360). The current study employs one particular kind of research synthesis—the statistical meta-analysis—to answer new questions about distance education instruction with previously obtained data (Mayer, 2010).

That reservoir of previously obtained data—the meta-analysis universe—provides ample material upon which to conduct a research synthesis. Many of the studies forming this universe are in direct contradiction with each other or involve dramatically different study populations. Both circumstances create difficulties for those looking for some point of consensus or mutual agreement. The statistical meta-analysis procedure is one of the most rigorous forms of research synthesis and is particularly adept at detecting overall patterns of cause and effect in diverse data collections.

A statistical meta-analysis, according to Glass (1976, 1978b), compares the results of individual studies by translating those results into a standardized metric he called "effect size." An effect size is a proportion that compares the differences between the mean of two sample distributions as measured in standard deviations. The two distributions can be either from a control group and a treatment group (also called an experimental group) or from the pre-treatment and post-treatment performances of the same group. By comparing the difference or change between the mean of the two groups in terms of standard deviations, the effect of the treatment on the experimental or post-treatment group can be estimated. The advantage to this statistical translation is that the resulting effect sizes can be used to compare studies that use different dependent measures. Effect size is calculated, according to Glass's (1976) original formula, as follows (Formula 1):

$$\text{Effect size} = \frac{\text{Experimental mean} - \text{Control mean}}{\text{Pooled standard deviation}} \qquad (1)$$

Mathematically this is expressed as (2):

$$\text{Hedge's } g = \frac{M_e - M_c}{sd_p} \qquad (2)$$

Where M_e = Mean of experimental group, M_c = Mean of control group, and sd_p = the pooled standard deviation. Hedge's *g* uses a complex correction formula to calculate the standard deviation which will be used in this study and is described in detail later in this chapter.

The following detailed description of the methodology used to analyze the above study population generally follows the American Psychological Association's Meta-Analysis Reporting Standards (MARS), as described in its sixth edition publication manual (APA, 2009), but is modified for use in a dissertation.

Study Procedure

Look for learning activities. Unlike the bulk of previous studies of distance education, this study is designed to look for significant differences across *learning activities* rather than across *media delivery types*, because the well-established non-effect for media effectively controls for media type. In other words, f2f and DE can be treated as equivalent as far as delivery is concerned—unless there is some effect other than media type that accrues from one or the other. For instance, it is possible that a heretofore undetected benefit is derived from physical proximity for f2f instruction or a similarly undetected benefit derived from certain media abilities (such as being able to revisit discussions, lectures, etc.) for DE instruction. It is beyond the scope of this study to investigate either of those possibilities or to control for them in the absence of such research. For the current study, it is assumed that any such effect—if it exists—contributes more or less equally to activities conducted via each medium and any difference between f2f and DE is likely to be irrelevant to the comparisons made here.

This study is also designed to detect interaction effects, that is, effects obtained only when DE is combined with some instructional methodology or activity. Such effects, if any, can be either beneficial or detrimental. Compound interaction effects may also exist; that is, three or more contextual factors may interact to provide an effect where any two do not. Those effects are very difficult to tease-out of original research studies and even more difficult to analyze in large, diverse bodies of research. The present study is not designed to detect complex interactions of the sort mentioned above.

Employ comparative meta-analysis. A comparative meta-analysis differs from a traditional meta-analysis in that the comparison is between more than two contrasts rather than between two contrasting groups. The comparison of treatments is made according to the magnitude of the effect size calculated for each group of studies addressing a particular treatment. The following outline provides an overview of the procedure followed in this study:

I. Part I

1. Calculate an overall main effect for Web-based distance education (i.e., all included studies combined).
2. Group studies into chronologically defined groups (i.e., studies published 2005 and earlier; studies published 2006–2008 and studies published in 2009 –2010).
3. Calculate an effect size for each chronological group
4. Compare the results to each other and to Means et al. (2009) and Bernard et al. (2009).

II. Part II.

1. Create groups according to IS, calculate an effect size for each coding category
2. Rank order the results
3. Create sub-category groups for IS; calculate an effect size for each
4. Rank order the results by IS category
5. Repeat steps 5 and 6 for CD
6. Rank order the results

Part One: extending Means and Bernard. The study population for Means et al. (2009) and Bernard et al. (2009) were sampled from the meta-analytic universe of studies that existed on July 31, 2005 and December 31, 2005 respectively and that met the criteria previously articulated. From those study samples were extracted all studies that featured Web-based delivery of distance education. That population was extended by adding to the original study sample used by Means et al. (2009) and Bernard et al. (2009), all available studies meeting the same criteria that existed on December 31, 2010, including some studies that were missed by the earlier meta-analyses. The new studies were combined with the original studies and main effects for Web-based delivery of distance education were calculated and compared to the main effects in the original study.

The main effects and contrasts for the current study are compared to the original main effects and contrasts as found in Means et al. It is hypothesized that no significant difference in main effects or contrasts will be found between the current, extended study and the original study, thus extending by another two years the historical finding of no significant difference between DE and f2f delivery of instruction.

In addition to the comparison for significant difference in average effect size between the earlier studies and the more recent studies, a time series comparison was contrasted between three chronological time periods: the period covered by Means et al. (2009) and Bernard et al. (2009) for 2005 and earlier, the time period during which an “explosion” of studies of Web-based distance education was published (2006–2008) and the most recent research published during 2009–2010 that reflected a rising awareness and utilization of Web 2.0 technologies in Web-based distance education. Effect sizes for each time period were calculated and the effects sizes compared.

Part Two: drilling down into Means et al. and Bernard et al. (2009). The second part of the current study is designed to drill down into the particular findings of Means et al. (and other recent meta-analyses) that ascertain that the majority of the variation between studies is attributable to two factors: time on task and equivalence of curriculum and instruction. The effect of time on task on student outcomes is well documented (see

Chapter 2) and it is unnecessary to further examine that factor to determine that extending time on task is part of best practices for Web-based distance education. Instead, this study focuses on the second factor, the role of differences in curriculum and instruction on the achievement of students in Web-based distance education. The three other meta-analyses identified various elements of this factor as having an effect: pedagogy (Bernard et al., 2004b), instructor involvement (Zhao et al., 2005), media involvement (Zhao et al., 2005), type of interaction (Zhao et al., 2005), and instructional methods (Sitzmann, et al., 2009). Means et al. found that the instructor (same or different) made little difference in student outcomes, but the instructional materials and approach used did. However, Means et al. only coded instructional materials and approach using four categories: identical, almost identical, somewhat different or different. These coding categories do not differentiate *how* the instructional materials or approach were different, only that they were. The present study drills-down into the instructional materials and approaches described in the included studies to identify the relative effects of four categories of IS and four categories of collaboration on student achievement.

Bernard et al. (2004b) also found that the quality of study methods affected the contrasts, a confound that Means et al. (2009) controlled for by limiting their meta-analytic population to only experimental and quasi-experimental studies. Study quality in the current study is addressed in the same way: All studies used to calculate effects sizes in this study were either experimental or quasi-experimental studies.

Statistical software. The software used for all statistical calculations in this meta-analysis was *Comprehensive Meta-Analysis* by BioStat, version 2.2 (BioStat, 2009). Comprehensive Meta-Analysis (CMA) is a commercial statistical meta-analysis software package that was developed with the medical field in mind and is the most popular meta-analysis software in that field. Consequently, it is also the most expensive, though the author was provided with a full licensed copy at a reduced price due to his graduate student standing.

According to Bax et al. (2007), their comparison of meta-analysis software programs showed that CMA had the highest Internet profile of the software studied, was accurate and “scored highest on usability and . . . also [had] the most complete set of analytical features” of all but one other program (p. 1).” One of the outstanding features of CMA is its ability to handle direct input of a variety of statistics and to perform automatic transformations or conversions of the statistics as necessary in order to combine the study-level statistics. The authors conclude:

> Comprehensive Meta-Analysis . . . distinguishes itself from other programs by the option to enter effect sizes of different formats and comprehensiveness of the numerical options and output. Data can be entered manually or via copy-and-paste in the CMA spreadsheet; direct import of text or other data files is not possible. The program features all major graphical presentations. The tutorial and manual are to-the-point and extensive. The program is actively maintained and the website is modern and regularly updated. (p. 11)

The latest version of CMA, used for this meta-analysis, can directly import data from Excel worksheets. This was particularly useful as coding of the studies was entered directly in Excel spreadsheets, thus eliminating the necessity of copy-and-pasting data a portion at a time.

Inclusion and Exclusion Criteria

The study population (or meta-analytic universe) for the present study was assembled using the criteria for inclusion as given by Means et al. (2009) with the following exceptions:

1) Studies (or effect sizes extracted from studies with more than one effect size) must be from Web-based distance instruction only. Non-web-based instructional situations such as Instructional Television or blended-class situations are not included. Web-based distance education courses are those where the entire content and all of the contact between teacher and student and between student and student occurs via the World Wide Web. It is recognized that in some cases students may contact each other in ways other than via the Web, but such contact is not officially a part of the course planned instructional activities.
2) To control for time on task effects, studies where the contrast featured a difference in treatment length were excluded.
3) Studies before July 2005 are limited to those included in Means et al. or Bernard et al. The assumption is that the researchers in those two studies have already identified all relevant studies and further searches of that material would result in few or no new studies being included.
4) Studies included in the present meta-analysis that were not included in Means et al. (2009) and Bernard et al. (2009) , used the same inclusion criteria as Means et

al. with the exception of those criteria presented here and which were published between August 2005 and December 2010.

5) Only studies that are readily available through the Lied Library services at the University of Nevada, Las Vegas or the Internet are included. Studies that might have been included but that were available only at an extra cost or were not obtainable through the Lied Library services were not considered nor included.

Data Collection

Data selection. Data selection proceeded in three parts: (1) locating the sample studies used in Means et al. (2009), (2) locating the sample studies used in Bernard et al. (2009) and (3) locating studies published since July 2005 using the criteria and search strategies described above. No systematic searches were performed for the time period prior to July 31, 2005 as it is assumed that the combination of Means et al. (2009) and Bernard et al. (2009), both with considerably more resources at their command than the present author, would have found all the relevant studies that were extant at that time.

(1) Studies from Means et al. (2009). Of the 51 studies coded in the Means meta-analysis (studies used by Means et al. only for their qualitative narrative analysis were not included), only four were not immediately available online through the UNLV Lied library online access. One article was only available as a paid article and three had to be located using resources other than the Lied library access. The remaining 47 studies were located using *Academic Search Premier*, *ERIC* and *Pro-Quest Dissertations and Theses*, and downloaded. Of those, 10 were found to meet the inclusion criteria for the present study. The primary differences between the present meta-analysis and Means et al. was in delivery method (only studies that included at least one group of subjects that used Web-based DE only were included, studies with solely non-Web delivery or blended instruction were excluded), subject type (only higher education or professional adult learners were included, all others were excluded) and outcome type (only studies or groups within studies that reported achievement outcomes were included, studies that reported only attitude, satisfaction or retention outcomes were excluded). In addition to the 10 studies from the meta-analysis sample, one Web-only based study not included in Means' meta-analysis but listed in their reference was found to include all the necessary information and data to be included in the present meta-analysis, bringing the total studies located from Means et al. (2009) to eleven.

(2) Studies from Bernard et al. (2009). Bernard et al. (2009) used 74 studies in their meta-analysis comparing DE instruction to other forms of DE instruction. Of the 74 studies used by Bernard et al. in their meta-analysis, 20 met the more rigorous criteria for inclusion in the present meta-analysis. The primary differences were: Bernard et al. included studies of any type of media delivery for DE whereas the present study was limited to DE delivered via the World Wide Web only. Bernard et al. included studies where some types of f2f meetings were included in addition to the DE portion; the present study employed a strict no f2f contact rule for inclusion. Bernard and associates included studies where subjects of all ages were included. The current study includes only higher education and professional adult learners. Bernard et al. also included studies that reported only attitudinal or retention data; the present study was limited to studies that reported at least one achievement outcome measure.

The present study was slightly more liberal in two inclusion criteria than were Bernard et al.: They excluded studies that reflected a treatment period of less than 15 hours and took place in a setting other than educational institutions. The present study did not exclude studies based on either criterion. Given the slightly more liberal inclusion criteria in the present study regarding treatment length and institutional setting than those in Bernard et al., searches for additional studies not included in either Means et al. (2009) or Bernard et al. (2009) extended back to August 1, 2005 in order to locate any study meeting those particular criteria that may have been published after July 31, 2005. The differences between the current meta-analysis and Means et al. and Bernard et al. are presented in Table 4.

Search strategies for additional studies. To locate new studies published since July 2005, databases available through the UNLV Lied library's online system were searched at least twice each between December 2009 and January 2011. The final search for each database occurred in January 2011 and included only those dates since the latest previously searched date for that particular database. In this way, studies published through December 2010 on each searched database were included in the present study.

The databases searched included the five used by Means et al.: *ERIC*, *PsychINFO*, *PubMed* (via Academic Search Premier), *ABI/INFORM*, and *UMI ProQuest Digital Dissertations*, but was extended to include Cambridge Abstracts. Test searches of Academic Search Premier were conducted to compare the return of titles using "Distance Education" as the keyword (that is, for all categories of search simultaneously) and those used by Means et al. found in Exhibits A-1 and A-2 on page A-2 of Means et al. (2009). Search engine technology has improved considerably in the past 4-5 years and varied search terms are less necessary today than they were just a few years ago. The test searches demonstrated that using "Distance Education" as a keyword for any category

Table 4

Differences in Inclusion Criteria

Means et al. (2009)

- Included all delivery media (f2f, blended, DE)
- Included non-achievement outcomes
- Included K-12 learners in addition to adults

Bernard et al. (2009)

- Included multiple DE media (ITV, etc)
- Included non-achievement outcomes
- Study quality was weighted, not controlled via inclusion criteria
- Treatment length was limited (none less than 15 days in length)
- Included subjects of all ages, not just adults
- Limited to formal educational institutions only

Current Study

- Web-based only
- Higher education or professional adults only
- Controlled for time-on-task between contrasts
- Added studies since July 2005

of searches produced more total returns than did the search strategy employed by Means et al. Using "Distance Education" was a less restrictive search strategy than that used by Means and meant that the returns were less targeted than in Means. The result was more time required on the part of the human searcher to filter the increased number of title returns. However, the benefit was an increased chance of locating potential studies for inclusion.

Some articles were only available in print and had to be located in the Lied Library stacks and photocopied. Most articles were available in digital format. In addition to the databases listed above, Google Scholar was searched using the same or similar terms as for the library database search (see Means et al., 2009). This resulted in relatively few potential studies that were not duplicates of those located using the library databases. Finally, the reference list for Sitzmann et al. (2009) was checked for references to any sample articles that had not already been located.

Criteria for document selection. The criteria for document selection followed those outlined in Means et al. (2009), with a few modifications. Briefly, articles were screened in two stages: an initial screening to determine if the study addressed online learning, if it used a controlled design and if it reported student outcomes. These procedures are described in more detail below. Studies that passed this initial screening were then examined more closely for four additional criteria for inclusion (adapted from Means et al., 2009, p. 12). A total of seven inclusion criteria were used in this study. A study was included if it:

1. Involved learning that took place via the World Wide Web (and associated Internet services) and involved adult learners;
2. Described an intervention that had been completed;

3. Compared contrast conditions, either equivalent control and treatment groups (between groups contrast) or repeated measure contrast within the same group (within group contrast), typically via the use of a pre-/post-test instrument;
4. Used a controlled design (experimental or quasi-experimental);
5. Reported an instructional intervention that provided the treatment to produce a contrast;
6. Reported an achievement outcome measured by continuous data; or
7. Reported data required to calculate or estimate an effect size (i.e., Hedges' *g*).

In respect to inclusion criterion 1 above, studies had to be solely online—blended studies or studies of treatment groups that met face-to-face at any point during the treatment time were not included. Further, online instruction was limited to Internet-based and delivered instruction. This is somewhat less restrictive than solely Web-based instruction, but reflects the fact that it is possible to conduct DE via non-Web technologies that use the Internet protocol and networking for delivery. Examples include e-mail, news groups, VoIP, RSS and other non-html-based technologies—even though many of those technologies are accessed via a Web-based interface. The distinction is a technical one, but a true distinction nonetheless.

Instruction that included media delivery other than via the Internet were included if that delivery was peripheral to—and supplementary to—the major instruction (that is, non-instructional). Examples of excluded studies that operated contrary to this criterion would be DE courses where the course materials were delivered on CD and print materials mailed through the postal system and the Web was used solely to maintain contact with the instructor rather than deliver instruction itself. In this case, while it is acknowledged that types of contact or interaction via communicative media can constitute instruction (depending, of course, on the content and purpose of the communication), the intent of the Web-based delivery criterion was to enforce particular delivery distinctions. Those distinctions include distinctions between blended instruction that uses Web-based delivery for a portion of the instruction (in which f2f interaction was the norm), correspondence DE in which multimedia-based instruction is delivered via the postal service (without the use of accompanying Internet-based communications; both of which involved physical artifacts exchanged between learner and instructor) and true digital DE where all contact between instructor and learner was in digital form sans any physical artifact or contact.

In respect to criterion 3, where the only between groups contrast was between f2f and DE groups, data were used only if the difference was between f2f and DE settings, where group baseline equivalency for the two groups was established AND where the groups differed in an identifiable instructional activity being delivered to the DE group. That is, f2f could be used only when the f2f group was used as a control group and when all instructional factors were the same except for physical proximity in f2f and some additional instructional intervention given to the DE group. In this case, the assumption, based on the previously discussed lack of demonstrable generic media influence or demonstrable physical proximity influence on learning, is that f2f and DE are equivalent as long as the instruction is equivalent. It is important to note that in such cases only the effect data of the Web-based group were included in the meta-analysis. The f2f groups were used solely as the contrast against which to derive the effect size.

In respect to criterion 6, studies had to provide objective measures of student learning. Instructor-graded items were included when a prior grading rubric or scheme was employed or when a standardized test was used that measured affective or attitudinal achievement, but required that the test be previously shown to be valid and reliable.

Multiple effect sizes from a single study. Multiple effect sizes from a single study were used if each effect size reflected a separate sample group, i.e., different subjects as opposed to simply different measures.

The ideal study for this meta-analysis. The ideal study for this meta-analysis is described below. Relatively few studies were located that could be considered "ideal."

For studies employing single-group repeated measures within group studies, the study would include the following:

- Employ some sort of pre-measure of the dependent variable to establish the baseline pre-treatment level of the dependent variable;
- Identify the instructional activity serving as treatment;
- Control as many confounds as possible;
- Measure post-treatment learning using an measure equivalent to the pre-measure;
- Report the descriptive statics for both pre and post measures that fully describe the distributions of each (at minimum: number of subjects, mean and standard deviation).

For studies employing between-group comparisons, included data should establish the following: (1) Establish the equivalency of groups, even when using a control group, especially when using convenience or self-selected samples, as is typical in educational research. (2) Limit treatments to one identifiable contrasting treatment per treatment group. (3) Measure post-treatment using the same measurement instrument for all groups, preferably one that is equivalent to the pre-measurement, if one is used.

When comparing f2f groups with DE groups, all instructional activities should be kept the same or equivalent except the contrast intervention/treatment. A common confound with f2f versus DE studies is the in-equivalency of the instructional interventions, leading to exactly the confound Mayer (1984) cautioned against.

Finally, all measures of learning should be measured using a ratio scale (continuous numbers). Statistical meta-analysis only works correctly when the outcomes being combined are composed of continuous number data. In the present case, the ideal study would also involve only Web-based samples where adults were the subjects—whether using a single group (repeated, within-group measures) or contrasting groups (between-group measures).

Inclusion procedure. Each title returned by database searches was initially scanned using very liberal criteria reflecting a philosophy of inclusion, that is, any excuse was accepted as a possibility for examining the abstract. A total of 7712 titles were returned by the various database searches and scanned for possible inclusion. Of those 7712 titles, 1373 (17.8%) included sufficient information to read their abstracts for further information. Abstracts were read for the presence of exclusionary criteria only. That is, the abstract had to explicitly include information that indicated that the study was not Web-based, did not involve higher education or adult learners, did not include measures of achievement, did not involve instruction as treatment or was designed as something other than an experiment or quasi-experiment. Any studies in which the abstract did not provide the information necessary to exclude them from the meta-analysis were downloaded for further inspection. Of the 1373 abstracts read, 511 articles (37.2%; 511/1373) were downloaded for detailed analysis. Of these, 88 were duplicates of studies included by Means or Bernard or from the database searches alone. Thus, the population of unique research articles from which the meta-analytic sample was to be comprised was 423 (30.8%; 423/1373).

The final stage of the inclusion/exclusion process involved two separate passes through each article. The first pass was exclusionary; that is, the same principles that applied to reading abstracts were applied to the entire article. Any article in which exclusionary data were found was excluded. The final step was to re-examine the article for the information and data necessary for inclusion. Of the 423 unique articles retained for detailed examination, 26 (6.1%) were retained for inclusion in the meta-analysis. The 26 studies retained for inclusion represented 0.34% (26/7712) of the original titles returned by searches of databases for the period July 1, 2005 through December 31, 2010.

An additional 13 titles were located from references in journal articles; based on the abstracts of those titles, ten articles were examined in greater detail. Of those ten, 3 met all the inclusion criteria and were included in the meta-analysis. When added to the 20 studies from those included in Bernard et al. (2009) and the 10 from Means et al. (2009), a total of 59 studies are included in this meta-analysis (see Table 5). Those 59 studies yielded 86 unique contrasts (i.e., effect sizes) and included a total of 5779 individual study participants. A complete list of the included studies appears in Table 6. Summary statistics for the search and inclusion process are detailed in Table 7.

Table 5

Sources of Included Studies

Source	Number	Percentage
Means et al. (2009)	10	17.0%
Bernard et al. (2009)	20	33.8%
Database Searches	26	44.1%
Other sources	3	5.1%
Total	59	100%

Descriptions of the Study Sample

The studies included in this meta-analysis range in date from 1998 through 2010, with every year during that time period, with the exception of 1999, represented by at least one study. The bulk of the studies from 2008 and earlier came from the studies included in Means et al. (2009) and Bernard et al. (2009). All but one of the studies included from 2010 were dissertations. The most productive period for studies that met all of the criteria for inclusion were the years 2006–2008, the period between the publishing of the first group of meta-analyses cited as influential (the last of those published in 2005) and the publication of the Means and Bernard meta-analyses. During that three-year period of time, 34 of the included studies were published—almost half (47.9%) of those included in the present meta-analysis. Only 13 of the included studies appeared in 2004 or earlier,

Table 6

Studies Included in the Meta-Analysis

Study Name	Source	Year	Journal Title
Adcock et al. 2006	C	2006	*The Quarterly Review of Distance Education*
Alavi, Marakas & Yoo 2002	B	2002	*Information Systems Research*
Anderton 2005	B	2005	Doctoral dissertation
Banks 2004	B	2004	Doctoral dissertation
Baturay & Bay 2010	C	2010	*Computers & Education*
Benjamin et al. 2008	M	2008	*Maternal and Child Health Journal*
Bernard & Lundgren 2001	B	2001	*Educational Research and Evaluation*
Bixler 2008	M, D	2008	Doctoral dissertation
Boulter 2010	D	2010	Doctoral dissertation
Caldwell 2006	B	2006	Doctoral dissertation
Castaneda 2008	M	2008	Doctoral dissertation
Cavus 2007	M	2007	*Journal of Educational Computing Research*
Chang & Chang 2008	C	2008	*The Quarterly Review of Distance Education*
Chen, B., Hirumi & Zhang 2007	C	2007	*The Quarterly Review of Distance Education*
Chen. C & Shaw 2006	B	2006	*Journal of Distance Education Technologies*
Clapano 2010	D	2010	Doctoral dissertation
Collins 2000	B	2000	*British Journal of Educational Technology*
Connolly et al. 2007	C	2007	*Computers & Education*
Cook et al. 2007	M	2007	*Medical Education*
Draper 2010	D	2010	Doctoral dissertation

Table 6 continued

Study Name	Source	Year	Journal Title
Fox 2010	D	2010	Doctoral dissertation
Frey 2008	C	2008	*Journal of Technology and Teacher Education*
Frith & Kee 2003	B	2003	*Journal of Nursing Education*
Gulikers, Bastiaens & Martens 2005	B	2005	*Computers in Human Behavior*
Gupta 2006	D	2006	Doctoral dissertation
Hairston 2007	M	2007	Doctoral dissertation
Hansen 2000	B	2000	Doctoral dissertation
Hansen 2008	C	2008	*Journal of Marketing Education*
Hylton 2006	C	2006	*Journal of Baccalaureate Social Work*
Isenberg 2010	D	2010	Doctoral dissertation
Jang et al. 2005	M	2005	*Journal of Nursing Education*
Jung et al. 2002	B	2002	Doctoral dissertation
Kanuka & Jugdev 2006	C	2006	*Open Learning*
Karatas & Simsek 2009	C	2009	*The Quarterly Review of Distance Education*
Karr et al. 2003	B	2003	*The Journal of Interactive Online Learning*
Kemper et al. 2006	M	2006	*BMC Medical Education*
Krall et al. 2009	C	2009	*Journal of Science Education and Technology*
LaRose, Gregg & Eastin 1998	M	1998	*Journal of Computer Mediated Communication*
Lee 2010	D	2010	Doctoral dissertation
Mebane et al. 2008	D	2008	*International Journal of Human-Computer Interaction*
Own 2006	C	2006	*International Journal of Science and Mathematics Education*
Pacifici et al. 2006	C	2006	*Children and Youth Services Review*
Parsons 2006	D	2006	Doctoral dissertation
Peterson & Bond 2004	M	2004	*Journal of Research on Technology in Education*

Table 6 continued

Study Name	Source	Year	Journal Title
Romanov & Nevgi 2006	B	2006	*International Journal of Medical Informatics*
Ruksauk 2000	B	2000	Doctoral dissertation
Schroeder 2006	B	2006	Doctoral dissertation
Seabolt 2008	D	2008	Doctoral dissertation
Sendag & Odabasi 2009	D	2009	*Computers & Education*
Shana 2009	D	2009	*Educational Technology & Society*
Skylar 2004	B	2004	Doctoral dissertation
Stanley 2006	B	2006	*The Journal of Educators Online*
Tsai, Tseng & Hwang 2008	D	2008	*International Journal of Distance Education Technologies*
Wallace et al. 2006	B	2006	*Journal of Interactive Learning and Research*
Westhuis, Oullette & Pfahler 2006	C	2006	*Advances in Social Work*
Williams 2005	B	2005	Doctoral dissertation
Wise et al. 2004	B	2004	*Journal of Educational Computing Research*
Yang, Newby & Bill 2008	C	2008	*Computers & Education*
Yavuz 2007	C	2007	*Turkish Online Journal of Distance Education*

Source Key: B – Bernard et al. (2009); C- Cambridge Abstracts; D – ABI/Inform/Pro-Quest and misc sources; M – Means et al. (2009)

Table 7

Detailed Statistics of the Search and Inclusion Process from Databases

Database	Titles Scanned	Abstracts Read	Articles Read	Articles Included
ABI/Inform/Pro-Quest	2166	361	116	7 (0.23%)
Academic Search Premier	1667	248	74	3 (0.18%)
Cambridge Scientific Abstracts	3879	764	311	16 (0.41%)
Misc sources	13	13	10	3 (23.1%)
Total from Databases	7725	1386	511	29 (0.38%)
Duplicates	NA	NA	88	-
Net totals	7725	1386	423	29
Percentage inclusion	-	17.7 %	30.52%	6.86%
		(1368/7725)	(423/1386)	(29/423)
	-	-	5.48%	2.1%
			(423/7725)	(29/1386)
	-	-	-	0.37%
				(29/7725)

with almost three-quarters (71.2%) of the studies published after the earliest rounds of meta-analyses effectively declared the media debate settled. The complete tabulation of the number of included studies by year of publication appears in Table 8.

Table 8

Number of Included Studies by Year of Publication

1998 – 1
1999 – 0
2000 – 3
2001 – 1
2002 – 2
2003 – 2
2004 – 4
2005 – 4
2006 – 15
2007 – 6
2008 – 10
2009 – 4
2010 – 7
Total = 59

Journals from the field of instructional technology and or distance education account for 21 of the articles, medical journals published six and social programming journals accounted for three more. Slightly more than one-third (21) of the studies included here are dissertations, including six in 2010 alone. Nearly half of the studies included in this meta-analysis came from diverse fields, some contributing only a single article. The articles were written and/or researched in locations all over the world, including four from Taiwan and three each from Canada and Turkey. The bulk, however, were written and researched in the United States, with Pennsylvania, Indiana and Texas leading all other states with four each, followed by Georgia and California with three each.

Study Sample Demographics

Demographic information about the included studies reveals that the academic fields of education, medicine and computer science account for the bulk of the studies included in the meta-analysis, but the fields of endeavor represented in the studies ranged from auto service supervisors and banking employees to foster parents. The bulk of the studies involved undergraduate students (66.1%) with the remainder involving graduate and professional career adults (see Table 9).

After identifying as many relevant studies for inclusion in the meta-analysis as logistically possible, the next step was to describe the sample studies and code them for use in the meta-analysis.

Table 9

Field and Standing of Subjects by Number of Studies

Field	Number	Percent	Standing	Number	Percent
Education	16	27.1%	Undergraduate	36	66.1%
Medicine[1]	10	17.0%	Adult/Career	12	20.3%
Computer Science[2]	8	13.6%	Graduate	11	18.6%
Business	7	11.9%			
General Education	4	6.7%			
Social Work	3	5.1%			
Science/Engineering	2	3.4%			
Psychology	2	3.4%			
Foreign Language	2	3.4%			
Environmental	2	3.4%			
Health & Safety	1	1.7%			
Career Training	2	3.4%			

[1] includes Veterinary Medicine

[2] includes Information Technology

Coding Procedures

Rationale for Coding Procedures

The coding list was intentionally kept simple for clarity. In addition to study identification information (citation, etc.), only information deemed essential to the purpose of the study were coded (Cooper & Hedges, 1994). The following information about each study was deemed essential to the meta-analysis and coded, where possible, for each study:

1. Effect Size (reported, calculated or translated)
2. Instructional Level: Undergraduate, Graduate, Adult/Career
3. Control/Treatment or Pre/Post-test design
4. Treatment Length
5. Demographic Information: subject matter area, treatment context (course) and the general location of the study
6. Assessment/instrument Type; student outcome(s) (i.e., dependent measure)
7. Research Question 1, lens 1: Instructional Strategy (IA) and sub-categories
8. Research Question 1, lens 2: Collaborative Design (CD)
9. Research Question 2: Chronological Group (CG)

Coding Categories

Examining the combined list of factors identified by Bernard et al. (2009), Zhao et al. (2005), and Sitzmann et al. (2009), reveals that all three meta-analyses identified activities by the instructor or the student as affecting student outcomes. Bernard, et al. (2009) drilled down into these moderator variables using three categories of interaction to identify efficacious DE instruction. This study attempts to drill even farther down by using four more focused lens' to code for specific behaviors involving instructors and students in order to identify those behaviors that are more effective at producing desired student outcomes than others. Those coding categories are described below.

Coding categories for research question 1. Means et al. (2009) cautioned that the beneficial effects sizes they reported for DE and blended instruction over f2f instruction were likely confounded by three moderating factors: in-equivalence of curriculum content, differences in pedagogy and differences in learning time. They noted that it may have been impossible to control for these factors in many instances. Bernard et al. (2009) also found that differences in pedagogy were important moderators of achievement outcomes in DE, but went further to identify three types of interaction that took place in an instructional setting: Student-student interaction, student-teacher interaction and student-content interaction.

The present study will focus on trying to determine what effect, if any, differences in pedagogy might make by coding every study in the meta-analysis sample using the following four lenses: (1) Instructional Strategy (IS)—that is, the over-reaching approach used by an instructor when designing and conducting instruction. There is no direct parallel to this in the interaction scheme used by Bernard et al. (2009), but it directly impacts all three interaction types identified by Bernard eta al. (2009); (2) Instructional Activities (IA)—that is, the specific instructional activities designed by the instructor that directly lead to student learning. This is analogous to the Student-Content interaction used by Bernard et al. (2009); (3) Instructor Role (IR)—that is, the basic instructional interaction(s) between Student and Teacher; and (4) Collaborative Design (CD)—that is, the arrangement and provision of collaborative activities available to and/or required of students. This includes both Student-Teacher interactions and Student-Student interaction. Each of these lenses is elaborated below, followed by a listing of the coding categories.

Instructional Strategy (IS). For the purposes of this research study, instructional strategy is defined as the over-reaching approach used by an instructor when designing and conducting instruction for higher education students. The design of the instruction for DE is the most significant predictor of increased achievement according to Bernard et al.

(2009); they found that “only strengthening SC interaction was related to increasing effect size (p. 1265).”

Originally used by Alexander Kapp in 1833, andragogy is a theory of adult education developed by the Malcolm Knowles (1980). It applies specifically to the individual subjects of this research study which limits itself to DE in higher education. Research in andragogy over the past two decades has focused on what is loosely termed “student-centered” education in contrast to what is equally loosely termed “traditional education.” Behind each of these terms stands an entire school of thinking on adult learning. According to Lepp (2010), student-centered instructional practices, as referred to in higher education, typically connote Constructivist educational theory. In like fashion, when people speak of “traditional instruction” they typically mean instructional practices associated with the Behaviorist theoretical position (Lepp, 2010). Both instructional theories are well-known to educators and need little introduction here.

Joyce, Weil and Showers (1992) suggest that all instructional methods and strategies can be categorized into four families of instruction: social, information processing, personal and behavioral systems. All—or most—instructional practices and methodologies fall into one of these families of instruction. Jackman and Swan (1996) used this instructional family model to classify individual instructional methodologies in their meta-analysis on effective instructional models for distance education. In their limited universe meta-analysis, they rank ordered numerous instructional strategies used with distance education using an interactive video system and found that the Role Playing, Simulation, Jurisprudential, Memorization and Synectics methods were more effective than other methods including Direct Instruction and Cooperative Learning. In a limited universe meta-analysis comparing different instructional strategies when using computers, Roberts (2002; 2008) found that using computers cooperatively and as Mindtools (Jonassen, 2000) were more effective than other methods studied.

Chickering and Gamson (1987) developed seven guidelines for instructional practice in undergraduate education primarily to help universities improve undergraduate education. As stated by the authors, these are the "teacher's *how*" rather than the typical "subject-matter *what*" [italics in original] (p. 4). Unlike the usual focus of instructor preparation for undergraduate instruction, these guidelines focus on what instructors should do (that is, the activities they engage in) rather than what content they should require students to learn. This is an important distinction because, as Archambault and Crippen (2009) discovered, K-12 teachers at least, tend to equate pedagogy with content.

They were not intended to be used as research categories, but literature on instructional competencies for post-secondary instructors is sparse (Chickering & Gamson, 1987, p. 4). Lou, Bernard and Abrami (2006), in a follow-up meta-analysis to Bernard et al., 2004), noted that their findings were consistent with Chickering and Gamson's seven principles (p. 1987). Originally, Chickering and Gamson's seven principles were planned as the basis for coding instructional practices for the current study. However, pilot coding tests demonstrated that insufficient detail existed in the bulk of the included studies to code at that level of detail.

Lepp (2010) investigated the level of knowledge about and use of non-traditional teaching practices by higher education instructors. In her study, Lepp differentiated between traditional or what she labeled “behaviorist” teaching practices (defined as lecture employed for about 80% of class time and students largely passive receptors of knowledge) and alternative or “constructivist” instructional practices, more commonly identified as student-centered instruction. Constructivist instruction focuses on student-centered learning activities where the students are active and collaborating with one another, the use of alternative assessments (in contrast to tests) and the use of modern digital computing technologies. She found that, while awareness of these instructional practices had increased compared to previous similar studies, the actual use of those practices had not increased commensurately. In contrast, she found that the use of digital computing technologies had risen significantly.

In contrast to the tendency for many higher education instructors to rely upon lecture as their primary instructional methodology (Becker & Watts, 2001; NCES, 2002), a practice that they derived from the way they were taught (Brown, 2003), research on student-centered instructional practices has generally found positive effects when used with adults (Anderson, 1988; Flint, 2004; Skinner, 2007; and West, Kahn & Nauta, 2007). Lepp (2010) defines student-centered instruction as follows:

> Student-centered instruction is defined as an approach to the practice of teaching which, based on the needs and strengths of the student, engages the student in the learning process, and provides the student an opportunity to be involved in the planning and delivery of material, assessment of the learning outcomes, and evaluation of the overall learning process. This active engagement uses the student’s natural abilities and curiosities as a springboard toward a more complete understanding of the material and higher-order thinking skills. (p. 2)

Several researchers have found that student-centered instructional practices, particularly in conjunction with digital computing technologies, improve student achievement (Choi & Johnson, 2005; Flint, 2004; McShannon et al., 2006). Constructivism has been associated with the use of computer technology, particularly in the ability of each to leverage the strengths of the other (i.e., Jonassen, 1996; Resnick, 1994; Thornburg, 1996), almost as long as computers have been in classrooms (Papert, 1980).

Behaviorist—or traditional—education has its advocates, as well. Well-regarded instructional designers such as Dick and Carey (1985), Gagné and Briggs (1979) and Romiszowski (1981) have developed behaviorist instruction to a fine art. Traditional, behaviorist instructional techniques were shown to be highly reliable for training large groups of learners during World War II (Jonassen & Land, 2000). Behaviorist instruction is fundamentally transmissive and, under the transmissive view of instruction, lecture is an exceedingly efficient instructional method.

It is not within the purview of this study to discuss the theoretical differences between these two varying points of view, but they do however, provide convenient categories into which studies with rather gross descriptions of their instructional practices can be placed. So, following Lepp, this study uses behaviorist and constructivist instructional strategies as the basis for coding and comparing instructional practices used in DE. While not as finely grained of a lens as would be preferable, identifying a DE course as being primarily behaviorist in design as opposed to constructivist subsumes all the instructional practices normally associated with that orientation. While not ideal, ascertaining whether there is any appreciable difference in the effect on achievement between the two instructional strategies would be informative in its own right.

Practice coding using just the two categories demonstrated that many, but not all, of the studies included in the meta-analysis sample could be identified as one or the other. In addition to obvious descriptions of lecture, discussion and readings, one determiner of the use of a behaviorist strategy was lack of designed student-student interaction and relatively sparse—typically student-initiated—student-teacher interaction. An additional identifying mark of behaviorist strategies was the lack of mass-individualized instruction. That is, what individual interactions that take place between student and teacher in a behaviorist instructional context are typically unplanned and occur as formative feedback and are either initiated by the individual student or involve the entire class as in "grading" of assignments or whole class discussions.

In contrast, constructivist instructional strategies are marked by a high degree of instructor-designed student-student interaction and more frequent individual student-teacher interaction. Typically, collaboration was a central instructional component of constructivist classes, but collaboration was coded separately and was not coded as a sub-category of constructivist instructional strategy.

Complicating the coding procedure was the frequent identification of specific instructional practices that were designed as contrasts within the individual studies themselves. Several frequently encountered instructional components of the respective strategies led to the creation of three sub categories; two for behaviorist and one for the constructivist. Sub-categories for Instructor Role (IR) and Instructional Activities (IA) were created to act as moderator variables within in the behaviorist strategy category. Each of these sub-categories, in turn, had several levels.

Four frequent instructional activities were identified as sub-categories for studies wherein constructivist instructional activities were identified as a contrast. These activities did not have subordinate levels. The four activities frequently mentioned as contrasts when constructivist instructional strategies were used: Simulations, Modeling, Concept Maps/Advanced Organizers and other (to keep the number of coding items for constructivist coded studies the same across all such studies).

In addition to the two major categories detailed above, two other frequent instructional strategies that did not fit either the behaviorist or the constructivist paradigms were identified. The first was the quite common media comparison study that had so dominated the early days of Web-based DE research. In these studies, because media delivery was the contrast, very little—if any—description of actual instructional design or activities was included in the study report. In these circumstances, it seemed reasonable to group those studies together. Thus, a third category of instructional strategy was created to complement the behaviorist and constructivist categories.

A fourth common occurrence were the instances where the description of the instructional activities was sufficient to identify an instructional contrast, but the interactions necessary to code for either behaviorist or constructivist strategies was lacking. In this case, the descriptions of instructional activity included elements of media delivery, but with a specific instructional activity designed as a contrast. What emerged from an examination of the descriptions was a situation that strongly featured learner control over the pace and sequencing of the material covered and no provision for student-student interaction and with student-teacher interaction limited to the student turning in assignments, the teacher grading and returning them and, occasionally, the student asking the teacher a question (which was invariably via some asynchronous communication system such as e-mail). The number of such studies led to the creation of a fourth category of instructional strategy called, appropriately, Independent Study.

No evaluation of the relative efficacy of any of the four learning strategies is implied here. Whatever differences there may be in their relative effectiveness is left to the statistical analysis to determine.

Collaborative Design (CD). The second major coding category is collaborative design. This differs from the categories of interaction used by Bernard et al., in that it solely deals with planned interactions between student-student, between student-teacher and between student-student-teacher. In this case, four mutually exclusive categories of planned collaboration/cooperation were formed to drill down into several observations made by Bernard et al. (2009) concerning SS and ST interactions. First, they noted that just because opportunities for student-student interaction were afforded, learning was not necessarily guaranteed (p. 1264). Second, they concluded that "courses lacking either mediated synchronous interaction or direct face-to-face interaction would benefit most from enhanced interactive capabilities (p. 1265)." Due to the structure of their coding scheme, Bernard and colleagues were unable to investigate these matters in greater depth. Coding for designed collaboration types in the present study will allow a more detailed investigation into the relative value of particular varieties of collaborative strategy. The four categories of collaborative design coded in this study are: none (no provision for collaboration); collaboration afforded, but not required; collaboration required and moderated by the teacher and collaboration required but left to students to facilitate the collaboration. Teacher moderation includes moderation of collaboration by student teaching assistants, even though it is possible that student teaching assistants may provide a different level and type of interactive presence than do course instructors.

Coding categories for research question 2. A somewhat different method needs to be employed to discern whether any progress has been made as a result of the continuous study of DE over the past decades. First, comparisons need to be made between various choices made in DE instruction rather than between DE and f2f; for instance, between different types of instructional strategies that all use Web-based technologies for the delivery of instruction. In the historical progress of DE, World Wide Web-based delivery of distance instruction using the Internet and digital computing technologies is the most advanced technology thus far employed for DE. The capabilities of current Web-based DE subsume all prior DE technologies. Thus, Web-based DE should logically be capable of delivering the most effective instruction used to date for DE. Following-up on that suggestion, the current study investigates whether any measureable progress has been made by comparing the effect sizes for Web-based DE conducted at different chronological points in the 13-year history of Web-based DE, using the same set of standards and the same metrics. Thus, the second research question addressed by this study is: Have Web-based distance education outcomes improved over time?

Second, because effects require contrasts and meta-analysis is capable of aggregating contrasts to detect otherwise small differences between contrasts, comparisons of the outcomes of DE instruction to detect increased effectiveness of that instruction using meta-analysis must be made between contrasting groups. In the present case, the assumption is made that instruction is more likely to increase in effectiveness over time than to decrease; contrasted groups were formed of temporally adjacent studies grouped according to a theoretical and practical scheme.

The included studies were grouped into three somewhat arbitrary *Chronological Groups* (CG), aggregate effects sizes for each of the groups were calculated and the results were compared to each other. The working hypothesis is that there will be a discernable, if not significant, difference in the effect size for the oldest group in comparison to the youngest group and an equally discerning, but less significant, difference between the middle-aged group and those on either side of it.

Chronological Group 1: 1998 – 2005. Prior to 1998, very few studies investigating Web-based DE were made, making it difficult to aggregate a reasonably-sized group for analysis earlier than that year. In 2005, Sitzman et al. and Zhao et al. published their meta-analyses, preceded by only few months by Bernard et al. 2004. Together, these three meta-analyses mark a maturing in the field of DE research. Each, in its own way, departed from the prevalent practice of purely media delivery comparisons and began looking at methodological issues, both in the conduct of the research studies used in their meta-analyses and in the instructional differences between studies. As suggested earlier, the field of DE research had moved on and it could reasonably be expected that their results would impact subsequent DE instruction.

Chronological Group 2: 2006 – 2008. This marks the period of time between the publication of the three studies mentioned in Group 1 above and the publication of meta-analyses by Means et al. (2009) and Bernard et al. (2009). Meaningful studies published during this time were not available when that first round of meta-analyses were published, but were available for inclusion in the Means and Bernard (2009) meta-analyses.

Chronological Group 3: 2009 – 2010. These are the studies published since Means et al. (2009) and Bernard et al. (2009) meta-analyses were published and analyzing these new studies extends by two years the body of evidence upon which any conclusions regarding the current effectiveness of DE can drawn.

The Coding Instrument

Below is an outline of the coding categories adopted for use in this meta-analysis. It is important to recognize that in a meta-analysis, the pool of subject studies is limited to those actually exist. The researcher cannot simply recruit more participants or conduct a follow-up study to compensate for missing data. The data is what it is and data that does not exist cannot be coded. Such is the case in the present study. As noted by Bernard et al. (2009) the body of appropriate and rigorous studies from which to build a meta-analysis sample for DE research is limited. Thus, some of the contrasts listed below as coding sub-categories could not be included in the analysis for the simple reason that insufficient studies exist to provide the data necessary for analysis. With that recognition, the categories used to code for the moderator variables analyzed to answer the research questions are presented below:

I. Instructional Strategy IS (ST Interaction):

1. <u>Media Delivery</u>. Studies identified as media delivery contrasts feature equivalent instruction and focus only on the difference in instructional delivery. These studies all feature contrasts between DE and f2f instruction and typically attempt to make instructional features as equivalent as possible. Thus, this category of instructional

strategy essentially codes studies with no instructional contrasts. The sub-categories of media delivery are:

A. non-multimedia – text-only delivery of curriculum materials

B. multimedia – one-way delivery of text, graphics and sound

C. interactive – media that allows for two-way interaction between parties

2. <u>Behaviorist (traditional) Instructional Strategy</u>. These are studies that involve nonequivalent, teacher-centered instructional contrasts where the primary contrast is some difference in teacher-centered instructional activity, irrespective of delivery differences. In all cases it is assumed that course design and evaluation (Summative feedback) are constant teacher activities. There are two sub-categories:

A. Instructor Role (IR)

   1. Motivation
   2. Formative feedback
   3. Deliverer of content/direct instruction only
   4. No active role described

B. Instructional Activities (IA)

   1. Simulations
   2. Modeling
   3. Case studies
   4. Concept maps/Advanced organizers
   5. Miscellaneous/other
   6. None/not described

3. Constructivist (student-centered) Instructional Strategy – These are studies that involve non-equivalent, student-centered instructional contrasts. The primary contrast is some difference in constructivist, student-centered instructional activity, irrespective of delivery differences. There are four sub-categories:

A. Inquiry Learning

B. Problem-Based Learning

C. Scaffolding

D. Other

4. Independent Study – These are studies that involve non-equivalent instructional strategies whose primary contrast is the presence of learner control, with Student-Teacher interaction restricted to assignment feedback or individual questions and no Student-Student interaction. This is always in contrast to Behaviorist f2f and the distinguishing feature is provision for learner control or self-pacing and absence of instructor sequencing and/or pacing. [No contrasts]

II. Collaborative Design CD (SS Interaction)

1. No provision for collaboration

2. Collaboration afforded (not required’ voluntary use)

3. Collaboration required, teacher moderated (evaluated)

4. Collaboration required, not moderated (i.e., student facilitated) Note: this includes both collaborative and cooperative learning SS types and includes both teacher evaluated and non-teacher evaluated. That is, some collaborative/ cooperative work is

directly evaluated as cooperative/collaborative work per se, as opposed to evaluation of some artifact that results from such collaboration.

III. Chronological Progress

1. Group 1: Studies published between 1998 and 2005 inclusive
2. Group 2: Studies published 2006 – 2008, inclusive.
3. Group 3: Studies published 2009 – 2010, inclusive.

Coding of Studies

Coding itself took place directly into the computer using Microsoft Excel®. A spreadsheet was created with categories for study characteristics (study number, citation, type of study), population characteristics (number of subjects, instructional level, demographics, selection basis–random or not), environmental characteristics (subject matter studied and length of treatment), and statistical characteristics (independent/ dependent variable, statistics reported for each). At this stage, coding was a relatively simple process of locating the relevant information and entering it in the appropriate spreadsheet location. All information was recorded, including information for multiple groups or studies within single studies. Particularly important at this point in the procedure was identification of the individual contrast groups in each study that met the criteria for inclusion as a separate effect size. Thus, some studies yielded two or more groups which independently yielded an effect size from a contrast meeting the criteria listed above.

Once the relevant data for each of the categories listed above was recorded, each contrast group was coded for categorical grouping according to two characteristics:

Instructional Strategy (IS) contrasted and Collaborative Design (CD) employed. Each included contrast group was coded for one IS and one CD. In addition, specific contrast information for IS was coded where described. The latter information was used to create sub-categories within each IS category that allowed for more in-depth analysis and exploration of confounds to take place after the main contrast effects were calculated.

Coding calibration. The categorization of IS and its associated sub-categories and of CD required some subjective judgment and was frequently open to alternative interpretations. Because of this and because the categories for comparison were central to comparisons being made in this study, it was necessary to standardize the coding as much as possible.

All coding was performed by the author. There were three primary reasons for this: First, the coding process required extensive background knowledge of instructional theory, strategy and techniques and an extensive familiarity with the terminology and jargon often used to describe instructional practices in research studies. It also required experience and familiarity with reading and analyzing research reports. Individuals with the requisite background were not available locally.

Second, due to the analytical nature of the coding process, it required the expenditure of a considerable amount of time. The costs of hiring, training and paying for the time required to code the studies were beyond the resources available and would have exceeded the time allotted for completion of the study.

Two volunteer pilot coders, one with background and experience in educational research and the other with background and experience in instructional design and practice, were used to test both the final coding categories and act as a comparison with the main coder. Each coder coded the same three studies and provided feedback. The author also coded the same three studies. Intercoder agreement for the pilot coding was calculated using the simple "joint possibility of agreement model" (see Miller & Vanni, 2005). While less robust than Cohen's κ, it is appropriate where the data being coded is primarily nominal in nature. Cohen's κ is designed to measure the reliability between only two raters, and other agreement schemes are not appropriate for nominal data. The "joint possibility" model ignores the possibility of chance causing agreement, but not all theorists believe that it is necessary to correct for chance, (cf., Uebersax, 1987). In this case, the type of coding being done closely modeled the Rasch model of inter-coder agreement which assumes that coders are independent witnesses to something that has happened and their independence is demonstrated by slight disagreement. Inter-coder reliability for the three pilot studies was 77.8%; with 83.3% agreement for Instructional Strategy (IS), but only 66.7% for Collaborative Design (CD).

The discrepancies in the initial coding were subsequently identified as due largely to the result of the distributed nature of the information presented in the study reports themselves, which led to pertinent information being missed by the volunteer coders. In follow-up discussion, the author brought the missing information to the attention of the volunteer coders at which point a consensus was reached. Feedback from the volunteers indicated that the coding guidelines were clear and sufficiently detailed to differentiate between categories and that no changes to the coding procedure or instrument were suggested. The purpose of inter-coder reliability studies, however, is to train consistent coders, and not necessarily to establish the reliability of the coding instrument. However, the pilot experience and discussion was used by the primary coder when coding the remainder of the studies.

Having a single coder, while eliminating differences of opinion as far as the actual coding is concerned, poses a different problem: the possibility of bias in the coding. Possible bias in coding in the current study was largely eliminated by the nature of the research questions: Without a hypothesis to prove or disprove, bias in favor of studies supporting the hypothesis was non-existent. The primary danger from bias in this case was the weighting in favor or one or more coding categories to the detriment of others. This danger was largely offset by the fact that the final coding categories were largely a reflection of the information and data contained in the included studies. Thus, bias in coding was directly tied to bias in selection—which is tested by multiple post hoc statistical tests. If bias in publication exists, then, it can be assumed that bias in coding would occur *ipso facto*.

Of greater danger than bias is error in coding. More than one coder provides a check on error, as discrepancies between coders can illuminate error and lead to timely correction. On the other hand, coding papers provides its own practice and practice helps produce proficiency. In the current study, each paper was fully coded twice, with a third partial coding focusing on the extraction of relevant statistical data. The second round of coding was particularly important for the studies that were initially the first to be coded, as the increased proficiency developed after coding all the sample studies resulted in numerous adjustments to the initial coded data. By recoding each study a second time, consistency was improved. Consequently, the chance that coding errors might influence the outcome of the meta-analysis is minimal.

To the extent that a single coder for this meta-analysis presents a potential for bias in the results, it is acknowledged that this is a limitation on the generalization of the study conclusions. The extent of that limitation, however, is tied to the extent of selection bias detected in post hoc statistical tests.

Identifying variables. Each study was represented by one or more identified Instructional Strategy (IS) as the independent variable (where available) and by student achievement as the dependent variable. In studies in which more than one contrast was reported, all contrasts involving separate groups of subjects with the requisite statistical and contextual data necessary to calculate either a between-groups or a within-group effect size were included. Where more than one instructional activity was reported as part of a treatment, the contrast most representative of the ontological basis of the study was used. When no other criteria led to a single identifiable effect size, the first reported contrast was arbitrarily chosen.

Statistical Methods

Reported Effect Size

Unless otherwise noted, this meta-analysis will calculate and report Hedges' *g* (the bias-corrected standardized mean difference) as the standard metric for effect size calculations. Hedge's *g* represents the bias-corrected standardized mean difference between the performance on assessments of learning by treatment or experimental groups and their associated control groups (between group measures). This applies also to the differences in pre-test and post-test scores by single groups (repeated measures; within group). The effect size is measured in standard deviations and indicates relative performance between the control and treatment (two or more groups) or between the pre-treatment and post-treatment scores on the same measure of achievement (one group). The meta-analytic procedure averages the relative performance of each study pretreatment group together and compares it to the average of the performance of each study post treatment group. The result is two distributions, one for the average performance of all the pre-treatment groups in the meta-analysis and one for the average performance of all the post-treatment groups from the same studies. These distributions normally overlap, but their means are separated by the difference of those means in standard deviations. In effect, the distribution curve of the post-treatment group is shifted, usually toward the right (i.e., higher or toward the hundredth percentile). This is illustrated in Figure 1 below using statistics from a previous meta-analysis that resulted in an effect size g = 1.03 of a standard deviation.

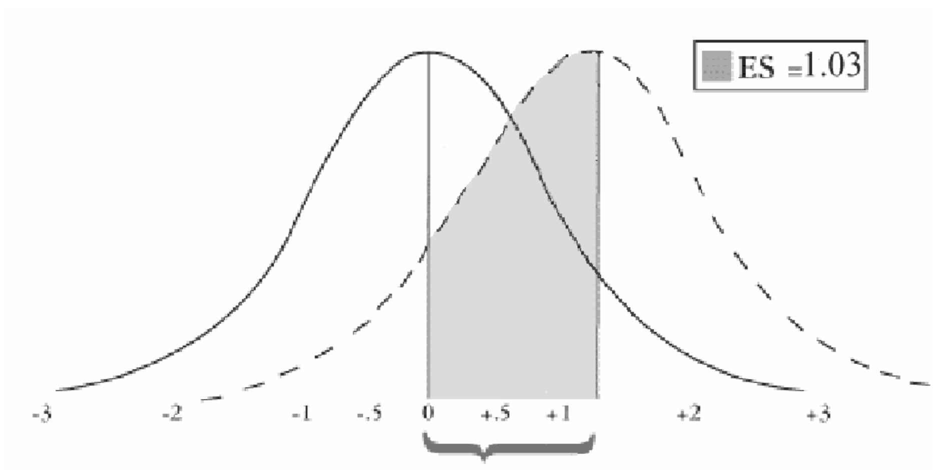

Figure 1. Graphical depiction of an overall mean effect of g = 1.03; modified from Marzano (1998).

Random Effects Model

Meta-analyses use weighted averages of the individual study results to generate an overall effect size (Egger, Smith & Phillips, 1997). The most typical procedures are those developed by Hunter, Schmidt and Jackson (1982) and Hedges and Olkin (1985). Two models exist to apply statistical techniques to this averaging process: The fixed effects (or conditional) model and the random effects (unconditional) model (Egger et al., 1997; Hedges, 1994a).

In meta-analysis, a random effects model uses both the within-study sampling error and the between-studies variation to generate the meta-analysis confidence level (Cochrane Collaboration, 2000; Raudenbush, 1994). The random effects model is used when sample populations and sample effects sizes are not homogeneous, that is, when "the observed variability in sample estimates of effect size is partly due to the variability in the underlying population parameters and partly due to the sampling error of the estimator about parameter value" (Hedges & Olkin, 1985, p. 191; Schwarzer, 1991, p. 29). The random effects model will be used in this study because a large difference in the study sample populations (and the subsequent effect sizes) exists, though fixed effects will also be calculated for use in post hoc analyses.

Missing Data

Studies with missing data necessary to perform the statistical analysis or without sufficient descriptive detail to code any of the lens categories were treated as ineligible studies—that is, they were eliminated from the study population.

Creating Study-level Summary Statistics

After all the relevant study information was coded and entered into an Excel spreadsheet, the next step was to convert each study statistic into a common measure (Egger, Smith, & Phillips, 1997; Lyons, 1998; Schwarzer, 1991). An effect size, Hedge's *g* statistic (unbiased effect size), was generated for each study by calculation from descriptive statistics or by transformation of *t* or *z* or by conversion from *F*. This creates the statistical effect of changing all the different fruits in the meta-analysis basket to apples so that apples could be compared to apples. Mixing apples and oranges has been a criticism of meta-analysis in the past and meta-analysts have typically dealt with that problem by using a number of corrections, weights, and controls to overcome it. By far the best way is to treat apples and oranges is to consider them as fruit where possible and compare them only in respect to their characteristics as fruit (Glass, 1978b; Glass, McGaw, & Smith, 1981; Rosenthal, 1990; Schwarzer, 1991; Smith, Glass, & Miller, 1980). The coding scheme used in this study creates, essentially, categories of "fruit" (IS or CD) from applicable contrasts.

Tests of Homogeneity

Homogeneity, the degree to which effect size estimates "exhibit greater variability than would be expected if their corresponding effect size parameters were identical" (Cooper & Hedges, 1994, p. 536) was calculated using the *Q* statistic (Cochran, 1954; Hedges & Olkin, 1985) using a .1 level of significance to detect hetereogeneity, I^2 to estimate the magnitude of any detected heterogeneity (Shadish & Haddock, 1994) and τ^2 to gauge the between-study variability. However, it is acknowledged that the sample used in this meta-analysis is highly heterogeneous by its very nature, so the normal use for heterogeneity tests in meta-analysis (i.e., to determine whether to use a fixed or a random effects model) is somewhat superfluous in the current study. The *Q* statistic is based on the χ^2 distribution and, while sensitive to the presence or absence of heterogeneity, does not quantify the amount of heterogeneity. The I^2, on the other hand, estimates the total percentage of variability among the studies due to true heterogeneity (Huendo-Median et al., 2006). *Q* is calculated as in (3):

$$Q = \sum w_i(T_i - T\text{-bar})^2 \tag{3}$$

The *Q* statistic, where w_i is the weighting factor for the *i*th study assuming a fixed-effects model, and *T* is defined in Formula 4.

$$\overline{T} = \frac{\sum_i w_i T_i}{\sum_i w_i}. \tag{4}$$

The final statistic describing heterogeneity is the τ^2 to estimate the between-studies variance. Under the random effects model, τ^2 indicates how much the true population effect sizes estimated in each of the individual study effect sizes differ from each other.

Corrections for Small Sample Bias and Unequal Sample Size

Because of the diverse nature of the studies that comprise the meta-analytic sample for this study, especially sample size and study rigor, each study was weighted according to sample size. Study rigor was controlled for through the adoption of inclusion criteria that weeded out all but the strongest studies. The assumption of heterogeneity in study size calls for employing a random effects model when combining the study level data, but following the suggestion of Mayer (2010), the main effect calculations were duplicated using a fixed effect model as a comparison. Typically, the drawback to using the random effects model is the presence of a slight bias in favor of small sample size studies, but this bias is eliminated or minimized through weighting each study according to sample size. The second control for using the random effects model is to run the data using the fixed effects model, which assumes homogeneous sample sizes and is more conservative in its results. Comparing the two outcomes provides a good test for the accuracy of the more appropriate random effects model.

Calculating Mean Effect Sizes

The final step was to calculate the mean effect size *g* for the entire study sample to derive an overall effect. Next, mean effect sizes were calculated for all sub-groups of the study sample grouped according to study date, instructional strategy and collaborative design. Post hoc tests for homogeneity and publication bias for each group and for the combined study sample were conducted simultaneously. Lastly, the Binomial Effect Size Display (BESD) (Rosenthal & Rubin, 1982) and the Common Language Effect Size (CLES) (McGraw & Wong, 1992) were calculated to aid in explaining what the effects sizes represented.

Tests for Data Censoring (i.e., Publication Bias, Selective Reporting)

Fail-safe N. Rosenthal (1984) described what he called the "file-drawer problem" which assumed that, in any given meta-analysis universe, an unknown number of non-significant studies with effect sizes of zero have either not been submitted for publication (reporting bias) or have been rejected (publication bias) and, so, have remained in file drawers somewhere. The fail-safe N calculates the number of these non-significant file-drawer studies required to bring the mean effect size down to a non-significant level as well. According to Rosenthal, it is possible to estimate the number of additional studies that would be required to reverse the overall *p* to a value higher than significance (Rosenthal, 1979, 1984, p. 108; Wolf, 1986, p. 38). The typical formula to estimate how many no-effect findings would have to exist in the file drawers in order to invalidate a significant overall *p* is shown in Formula 5.

$$N_{fs\ .05} = (\sum Z_i \ / \ 1.645)^2 - k \qquad (5)$$

Orwin's variant Fail-safe N. While Rosenthal's Fail-safe N assumes that unpublished studies have a nil effect (i.e., support for the null hypothesis), but it is possible that some unpublished studies do, in fact, show an effect—including some that may have a small effect in the opposite direction from the main effect. CMA also calculates Orwin's variant of the Fail-safe N to take into account studies that show a small negative effect. This statistic will be included in the analysis as an extremely conservative measure of the robustness of the study.

Plot by Precision. To test for publication bias based on the size of the study, a traditional funnel plot depicting the distribution of studies by Precision (calculated as 1/Standard Error) and the Log Odds ratio was generated. According to the CMA Manual (2005),

> In the absence of publication bias the studies will be distributed symmetrically about the combined effect size. By contrast, in the presence of bias, the bottom of the plot would tend to show a higher concentration of studies on one side of the mean than the other. This would reflect the fact that smaller studies (which appear toward the bottom) are more likely to be published if they have larger than average effects, which makes them more likely to meet the criterion for statistical significance. (p. 95)

Study Quality Issues

Meta-analyses are very sensitive to the quality of the individual studies that comprise the study population. In their meta-analysis of treatment effectiveness research, Wilson and Lipsey (2001) noted that "study methods accounted for nearly as much variability in study outcomes as characteristics of the interventions (p.413)." This variability due to quality is magnified when studies are combined statistically. To avoid—as much as possible—compromising the results of the present meta-analysis because of study quality only experimental or quasi-experimental study designs were included in the study population.

Missing studies. Both Rosenthal's Fail-safe N and Orwin's variant act as proxy indicators of the statistical power of the analysis. If the fail-safe N is large, then it can be assumed that the analysis study population included the most relevant studies and that any actual missing studies are unlikely to have a major effect on the average effect size. The funnel plot by precision (see above) and Duval and Tweedie's (2000) trim-and-fill procedure can act as proxy indicators of the statistical power of the meta-analysis. The funnel plot will provide some indication of the relevancy of the included studies and the trim-and-fill procedure will provide an approximate measure of the magnitude of any error in the analysis due to missing studies.

Large sample size bias. Large studies tend to be included in analyses regardless of their treatment effect whereas small studies are more likely to be included when they show a relatively large treatment effect. Under these circumstances, there will be an inverse correlation between study size and effect size. The Funnel Plot by Precision test (Egger et al., 1997) will indicate whether this occurred for most situations. If the funnel plot shows asymmetry then bias due to sample size may exist. To check whether asymmetry in the funnel plot by precision is due to bias caused by the size of the studies, the Begg and Mazumdar (1994) Rank Correlation Test (BMRC) is used. The BMRC computes Kendall's Tau-b, the rank order correlation between the treatment effect and the standard error, which is largely dependent upon sample size. If tau-b = 0, then no relativity exists and deviation indicates the presence of a relationship. If the asymmetry is caused by publication bias due to sample size, high standard errors (indicative of small sample size) will be associated with larger effect sizes. Egger's Linear Regression Intercept Method (ELRI) (Egger et al., 1997) uses the actual values of the effect sizes and their precision to quantify bias demonstrated in the funnel plot. It computes the standardized effect (i.e., effect size divided by the standard error). Comprehensive Meta-Analysis computes both the BMRC and the ELRI automatically whenever a funnel plot by precision is generated for a meta-analytic sample population.

Reporting Meta-Analysis Results

Reports for each included study contain the study citation and the sample statistics *n*, *g*, SE, variance, Lower Limit, Upper limit, Z and *p*. Study level statistics are reported in

Table 10. Mean effect sizes for groups are reported in Hedge's *g*, along with the study sample size *k* (number of effect sizes), subject sample size N (aggregate number of individual subjects in the sample studies), homogeneity χ^2, degrees of freedom *df*, significance levels *p*, and fail-safe N. Table 11, *Notation and Symbols*, lists the statistical symbols used in this meta-analysis.

Interpreting Meta-analysis Results

Two statistics to assist in interpreting the reported results were calculated and reported in this meta-analysis. The Binomial Effect Size Display (BESD) and the Common Language Effect Size (CLES).

Binomial Effect Size Display. Rosenthal and Rubin (1982) created a metric to statistically illustrate the movement described above called the Binomial Effect Size Display (BESD). BESD is the difference between the success rates of the post-treatment group and the pre-treatment group. It is calculated as in (6), where ES_t = treatment effect

$$(ES_t /2 + .50) - (ES_c /2 - .50) \tag{6}$$

size and ES_c = control effect size (Rosenthal & Rubin originally calculated this using *r*, but the principle holds for *g* as well.) The BESD will be reported as an additional explanatory metric in addition to the effect size.

Table 10

Study Level Statistics

Study Name	*n*	*g*	SE	Variance	Lower limit	Upper limit	Z	*p*
Adcock et al. 2006 1	71	1.400	0.186	0.035	1.034	1.765	7.507	0.000
Adcock et al. 2006 2	59	0.684	0.188	0.035	0.315	1.053	3.636	0.000
Alavi, Marakas & Yoo 2002	206	0.588	0.100	0.010	0.391	0.785	5.855	0.000
Anderton 2005	11	0.152	0.390	0.152	-0.613	0.916	0.388	0.698
Banks 2004 1	47	0.984	0.217	0.047	0.559	1.409	4.539	0.000
Banks 2004 2	25	0.927	0.293	0.086	0.352	1.502	3.159	0.002
Baturay & Bay 2010 1	35	0.108	0.237	0.056	-0.355	0.572	0.458	0.647
Baturay & Bay 2010 2	43	0.487	0.217	0.047	0.062	0.912	2.244	0.025
Benjamin et al. 2008	17	2.316	0.437	0.191	1.459	3.173	5.299	0.000
Bernard & Lundgren 2001	45	0.704	0.303	0.092	0.110	1.297	2.325	0.020
Bixler 2008	39	1.189	0.242	0.059	0.714	1.663	4.911	0.000
Boulter 2010	6	0.330	0.537	0.289	-0.723	1.383	0.614	0.539
Caldwell 2006	5	0.482	0.315	0.099	-0.134	1.099	1.533	0.125
Castaneda 2008 1	20	1.552	0.250	0.063	1.062	2.043	6.207	0.000
Castaneda 2008 2	41	0.989	0.213	0.045	0.573	1.406	4.655	0.000
Castaneda 2008 3	49	1.253	0.224	0.050	0.814	1.692	5.593	0.000
Castaneda 2008 4	47	0.426	0.214	0.046	0.007	0.845	1.994	0.046
Castaneda 2008 5	44	0.224	0.210	0.044	-0.186	0.635	1.071	0.284
Cavus 2007 1	45	0.053	0.326	0.106	-0.586	0.691	0.161	0.872
Cavus 2007 2	18	0.047	0.326	0.106	-0.592	0.686	0.145	0.885
Chang & Chang 2008	18	-1.091	0.219	0.048	-1.520	-0.662	-4.979	0.000
Chen, B., Hirumi & Zhang 2007 1	40	0.674	0.312	0.097	0.064	1.285	2.164	0.030

Table 10 continued

Study Name	n	g	SE	Variance	Lower limit	Upper limit	Z	p
Chen, B., Hirumi & Zhang 2007 2	21	0.179	0.303	0.092	-0.416	0.773	0.589	0.556
Chen, B., Hirumi & Zhang 2007 3	21	0.321	0.305	0.093	-0.276	0.919	1.054	0.292
Chen. C & Shaw 2006 1	21	0.279	0.240	0.058	-0.192	0.749	1.162	0.245
Chen. C & Shaw 2006 2	40	0.187	0.265	0.070	-0.332	0.706	0.705	0.481
Clapano 2010	26	-0.006	0.198	0.039	-0.395	0.383	-0.031	0.976
Collins 2000	50	0.107	0.227	0.052	-0.338	0.553	0.472	0.637
Connolly et al. 2007 1	22	0.359	0.316	0.100	-0.260	0.978	1.136	0.256
Connolly et al. 2007 2	14	0.945	0.249	0.062	0.456	1.434	3.788	0.000
Cook et al. 2007	19	0.000	0.215	0.046	-0.421	0.421	0.000	1.000
Draper 2010 1	57	0.901	0.189	0.036	0.531	1.271	4.768	0.000
Draper 2010 2	61	2.531	0.242	0.059	2.057	3.006	10.453	0.000
Fox 2010 1	61	0.121	0.222	0.049	-0.313	0.556	0.548	0.584
Fox 2010 2	40	0.335	0.223	0.050	-0.102	0.772	1.503	0.133
Fox 2010 3	40	0.025	0.221	0.049	-0.409	0.459	0.114	0.909
Frey 2008	40	1.371	0.459	0.211	0.471	2.271	2.985	0.003
Frith & Kee 2003	11	0.105	0.229	0.053	-0.344	0.554	0.458	0.647
Gulikers, Bastiaens & Martens 2005	40	0.079	0.335	0.112	-0.578	0.735	0.234	0.815
Gupta 2006 1	17	0.615	0.135	0.018	0.351	0.879	4.570	0.000
Gupta 2006 2	117	0.195	0.143	0.021	-0.086	0.476	1.359	0.174
Gupta 2006 3	85	0.541	0.133	0.018	0.280	0.801	4.064	0.000
Hairston 2007	120	0.672	0.167	0.028	0.345	0.999	4.023	0.000
Hansen 2000	75	0.296	0.110	0.012	0.081	0.511	2.699	0.007
Hansen 2008 1	188	2.064	0.444	0.197	1.194	2.934	4.648	0.000
Hansen 2008 2	15	1.676	0.319	0.102	1.051	2.301	5.257	0.000
Hansen 2008 3	26	1.823	0.380	0.145	1.078	2.569	4.795	0.000

Table 10 continued

Study Name	*n*	*g*	SE	Variance	Lower limit	Upper limit	Z	*p*
Hylton 2006	19	0.298	0.248	0.061	-0.187	0.784	1.205	0.228
Isenberg 2010	27	1.653	0.402	0.161	0.865	2.441	4.114	0.000
Jang et al. 2005	16	5.079	0.395	0.156	4.305	5.853	12.862	0.000
Jung et al. 2002 1	54	0.823	0.214	0.046	0.402	1.243	3.836	0.000
Jung et al. 2002 2	45	0.293	0.239	0.057	-0.176	0.762	1.226	0.220
Kanuka & Jugdev 2006	27	1.767	0.165	0.027	1.444	2.089	10.730	0.000
Karatas & Simsek 2009	102	1.913	0.309	0.095	1.307	2.518	6.191	0.000
Karr et al. 2003	30	0.386	0.321	0.103	-0.242	1.014	1.205	0.228
Kemper et al. 2006	14	0.079	0.051	0.003	-0.020	0.178	1.560	0.119
Krall et al. 2009	780	1.257	0.234	0.055	0.798	1.716	5.366	0.000
LaRose, Gregg & Eastin 1998	43	0.235	0.282	0.080	-0.318	0.788	0.834	0.404
Lee 2010 1	75	0.753	0.178	0.032	0.405	1.102	4.235	0.000
Lee 2010 2	67	0.817	0.169	0.029	0.485	1.148	4.827	0.000
Mebane et al. 2008	75	1.398	0.339	0.115	0.733	2.062	4.123	0.000
Own 2006	21	0.631	0.169	0.028	0.300	0.961	3.737	0.000
Pacifici et al. 2006	73	0.744	0.208	0.043	0.336	1.153	3.570	0.000
Parsons 2006	48	1.674	0.238	0.057	1.207	2.141	7.025	0.000
Peterson & Bond 2004	47	0.786	0.236	0.056	0.323	1.248	3.331	0.001
Romanov & Nevgi 2006	38	0.370	0.218	0.047	-0.056	0.796	1.700	0.089
Ruksauk 2000	39	0.166	0.143	0.020	-0.114	0.447	1.164	0.244
Schroeder 2006 1	96	0.719	0.263	0.069	0.203	1.235	2.732	0.006
Schroeder 2006 2	30	1.456	0.287	0.083	0.893	2.019	5.066	0.000
Seabolt 2008	30	1.083	0.333	0.111	0.431	1.735	3.254	0.001
Sendag & Odabasi 2009 1	20	1.429	0.349	0.122	0.746	2.113	4.099	0.000
Sendag & Odabasi 2009 2	20	0.553	0.316	0.100	-0.067	1.172	1.749	0.080

Table 10 continued

Study Name	*n*	*g*	SE	Variance	Lower limit	Upper limit	Z	*p*
Shana 2009	20	1.772	0.423	0.179	0.944	2.600	4.194	0.000
Skylar 2004	15	0.163	0.279	0.078	-0.384	0.709	0.583	0.560
Stanley 2006	25	0.403	0.398	0.159	-0.378	1.184	1.011	0.312
Tsai, Tseng & Hwang 2008 1	12	2.941	0.282	0.079	2.388	3.493	10.431	0.000
Tsai, Tseng & Hwang 2008 2	10	3.303	0.301	0.090	2.714	3.892	10.989	0.000
Wallace et al. 2006	52	0.686	0.302	0.091	0.093	1.278	2.268	0.023
Westhuis, Oullette & Pfahler 2006	52	0.885	0.281	0.079	0.334	1.437	3.146	0.002
Williams 2005	24	0.006	0.126	0.016	-0.241	0.254	0.050	0.960
Wise et al. 2004	27	0.228	0.430	0.185	-0.614	1.071	0.531	0.595
Yang, Newby & Bill 2008 1	98	0.323	0.413	0.171	-0.486	1.133	0.783	0.434
Yang, Newby & Bill 2008 2	10	0.598	0.587	0.344	-0.552	1.748	1.020	0.308
Yavuz 2007	11	-3.163	0.702	0.493	-4.539	-1.787	-4.506	0.000

Table 11

Notation and Symbols

<u>Symbol</u> <u>Definition</u>

BESD = Binomial Effect Size Display

d = effect size; Cohen’s standardized mean difference

đ = “d – hat;” mean effect size

ES = average effect size across a set of effect sizes; Glass's symbol

es = effect size for a single study

g = Hedge’s bias-corrected standardized mean difference

k = number of effect sizes

M = mean

N = number of subjects per study

N_{fs} = Fail-safe N

n = number of samples in a group yielding an effect size

p = significance

SD = average standard deviation across a set of studies

sd = standard deviation for a single study

χ^2 = Chi-squared; result of test of homogeneity

Common Language Effect Size (CLES). The Common Language Effect Size is a metric developed by McGraw and Wong (1992) to make it easier for non-professionals to interpret effect sizes. Essentially, the statistic gives the probability that a randomly selected score from the treatment group will be greater than a randomly sampled score from the comparison group. The CLES is computed by converting the effect size to a Z score and finding the probability of that score being greater than 0 (the mean).

Limitations of the Study

The following limitations exist in this study:

1) A limited study population was used: Only studies completed since 1997, and only documents located as of January 21, 2010 were included and only experiments and quasi-experiments studies were considered. Only databases available through the UNLV Lied Library or publically available on the World Wide Web were searched.
2) The reliability and validity of individual studies were not established.
3) Only one coder was used for the bulk of the studies.
4) A formal statistical accounting for the heterogeneous nature of some of the meta-analysis subgroups was not conducted, leaving the percentage of variation observed in effect sizes not accounted for by sampling error unaccounted for.

CHAPTER 4

RESULTS

Chapter Four presents the results of the statistical analysis and the findings of the study based on those results. This chapter includes coding results, grouping descriptives, modifier investigation and analysis and assessments of bias. The chapter also includes tables and charts illustrating the results of various post hoc tests. It concludes with a summary statement of the findings of this study.

Restatement of the Research Questions

This study seeks to answer two research questions in regard to the use of Web-based distance education. The two questions are re-stated here:

Research Question 1: Which instructional interventions are most effective when used in a Web-based distance education setting—and under what circumstances?

Research Question 2: Have Web-based distance education outcomes improved over time?

The Research Plan

To answer those questions, a statistical meta-analysis was used, following the general procedure listed below:

I. Part I

1. Calculate an overall main effect for Web-based distance education (i.e., all included studies combined).

2. Compare the results to Means 2009 and Bernard et al 2009.
3. Group studies into chronologically defined groups (i.e., studies published 2005 and earlier; studies published 2006-2008 and studies published in 2009 – 2010).
4. Calculate an effect size for each chronological group
5. Compare the results to each other to answer Research Question 2.

II. Part II.

6. Create groups according to IS, calculate an effect size for each coding category
7. Rank order the results
8. Create sub-category groups for IS; calculate an effect size for each
9. Rank order the results by IS category
10. Repeat steps 5 and 6 for CD
11. Rank order the results

Part 1: Main Effect and Research Question 2

The Main Effect and Its Comparison to Previous Meta-analyses

Using the study level statistics presented in Chapter 3 (pages 117), a main effect was found for the combination of all 86 effect sizes derived from the 59 studies that qualified for inclusion in this meta-analysis. The combined studies represented 5779 individual study participants and were selected for inclusion from a potential study population of 7725 titles that met the original database search parameters. As stated earlier, the random effects model was used for calculating effect sizes, but fixed effects effect sizes were also calculated for potential use as a check on the robustness of the statistical findings.

Main effect. The effect size is reported in Hedge's *g*—the corrected, standardized difference in means. The main effect for the combined studies is $g = .777$ ($k = 59$, SE $= .078$). The fixed effects model effect size and additional statistical data concerning the main effects are presented in Table 12 below. Both random and fixed main effects are significant, though there is considerable difference between the two. The fact that both fixed and random effects models are significant reinforces the robustness of the significance, though it should be noted that the fixed effects model is not statistically appropriate for this sample. Comparing the two, however, provides some insight into the nature of the sample characteristics. For example, the larger size of the random effects model effect size (compared to the fixed effects model effects size) typically indicates the presence of many strong, positive effect sizes from smaller studies—a possible publication bias that will require addressing later. For comparison, Table 13 presents the main effect using the common, but somewhat less rigorous measure, Cohen's *d*. The two effect sizes, using both fixed and random effects models, are almost exactly alike, varying by only one one-hundredth in all measures except for the Z scores, where they vary by about 15 one-hundredths. Note that regardless of which metric and which model is used, the main effect is significant. That is, on average, the difference in outcomes of the interventions reported in the sample of studies in this meta-analysis is significantly different (in a positive direction) than the contrast condition (i.e., either a within-group pre-treatment condition or a between-groups equivalent control group).

Table 12

Overall Main Effects of Web-based Distance Education (Hedge's g)

Model	Hedge's *g*	SE	Variance	Lower limit	Upper limit	Z	*p*
Fixed	0.553	0.022	0.001	0.509	0.597	24.653	0.000
Random	0.777	0.078	0.006	0.624	0.930	9.945	0.000

Table 13

Overall Main Effects of Web-based Distance Education (Cohen's d)

Model	Cohen's *d*	SE	Variance	Lower limit	Upper limit	Z	*p*
Fixed	0.552	0.022	0.001	0.508	0.596	24.582	0.000
Random	0.779	0.079	0.006	0.624	0.933	9.894	0.000

This concludes the first step in part one. The next step is to directly address Research Question 2 by generating effects sizes for groups created according to year of publication.

Research Question 2: Comparison of Group Effect by Year of Publication

Research Question 2 asked: Have Web-based distance education outcomes improved over time? To answer this research question, the included studies were divided into three groups according to year of publication. Group 1 included all studies published in 1998 through 2005. Group 2 included all studies published in 2006, 2007 and 2008. Group 3 included all studies published in 2009 and 2010. The studies included in each group are shown in Table 14.

Table 15 presents the results of the calculation of mean effect size for each of the three chronological groups. The aggregate effect size for Group 1 was g = .606 (k = 19, SE = .153). The aggregate effect size for Group 2 was g = .824 (k = 49, SE = .109). The aggregate effect size for Group 3 was g = .830 (k = 18, SE = .172). As shown in Table 16 below, all three groups are very heterogeneous, but all approximately equally so. There is less than a 4% difference in heterogeneity between all three groups (I^2 = 87.698 vs. I^2 = 90.038 and I^2 = 91.650). The results would seem to indicate that the effectiveness of Web-based DE has increased over the past 13 years.

Though the comparison is not exactly fair, it is worth noting that Zhao et al. (2005) recorded a combined effect size for DE studies published prior to 1998. Using Cohen's *d*, they found that d = - 0.10 (k = 20, SE = .11, Q = 24.400). Considering that Cohen's *d* is slightly larger than the commensurate Hedge's *g*, the difference in effect sizes between studies published prior to 1998 those published in 2009 and 2010 is dramatic. Thus, the answer to research question 2 appears to be "Yes, Web-based distance education outcomes have improved over time."

Conclusion to Part 1. This concludes the presentation of results for Part 1 of the study. In Part 2, the central focus of the study is addressed: As restated from Chapter 1, the purpose of this study is to identify some of the best instructional practices when distance education is the delivery method for higher educational instruction. It seeks to provide a preliminary identification of the instructional practices and methods that appear

Table 14

Composition of Groups by Year

Group 1: 1998-2005 (*k* =19)	Group 2: 2006-2008 (k =49)	Group 1: 2009-2010 (*k* = 18)
Alavi, Marakas & Yoo 2002	Adcock et al 2006	Baturay & Bay 2010
Anderton 2005	Benjamin et al 2008	Boulter 2010
Banks 2004	Bixler 2008	Clapano 2010
Bernard & Lundgren 2001	Caldwell 2006	Draper 2010
Collins 2000	Castaneda 2008	Fox 2010
Frith & Kee 2003	Cavus 2007	Isenberg 2010
Gulikers, Bastiaens & Martens 2005	Chang & Chang 2008	Karatas & Simsek 2009
Hansen 2000	Chen, B., Hirumi & Zhang 2007	Krall et al 2009
Jang et al 2005	Chen. C & Shaw 2006	Lee 2010
Jung et al 2002	Connolly et al 2007	Sendag & Odabasi 2009
Karr et al 2003	Cook et al 2007	Shana 2009
LaRose, Gregg & Eastin 1998	Frey 2008	
Peterson & Bond 2004	Gupta 2006	
Ruksauk 2000	Hairston 2007	
Skylar 2004	Hansen 2008	
Williams 2005	Hylton 2006	
Wise et al 2004	Kanuka & Jugdev 2006	
	Kemper et al 2006	
	Mebane et al 2008	
	Own 2006	
	Pacifici et al 2006	
	Parsons 2006	
	Romanov & Nevgi 2006	
	Schroeder 2006	
	Seabolt 2008	
	Stanley 2006	
	Tsai, Tseng & Hwang 2008	
	Wallace et al 2006	
	Westhuis, Oullette & Pfahler 2006	
	Yang, Newby & Bill 2008	
	Yavuz 2007	

Note: Only studies are shown, not effect sizes; some studies have more than one effect size. Thus, the difference between the k (number of effect sizes) for each group and the number of studies that appear in the list for each group.

Table 15

Comparison of Groups by Year: Main Effect by Groups

Group	Years	*k*	*g*	SE
1	1998-2005	19	.606	.153
2	2006-2008	49	.824	.109
3	2009-2010	18	.830	.172

Table 16

Tests for Heterogeneity within Year Groups

Year Group 1: Q = 180.690 (df (Q) = 18); I^2 = 90.038; τ^2 = 0.373 (SE = 0.184; variance = 3.396)

Year Group 2: Q = 574.859 (df (Q) = 48); I^2 = 91.650; τ^2 = 0.498 (SE = 0.209, variance = 0.044)

Year Group 3: Q =138.191 (df (Q) = 17); I^2 = 87.698; τ^2 = 0.443 (SE = 0.193, variance = 0.037)

to be more effective when used in a higher educational distance education setting by re-examining existing research using a type of research synthesis known as statistical meta-analysis.

Part 2: Answering Research Question 1

Coding Categories and Levels

To answer research question 1, studies were coded into two major categories of instructional activity: Instructional Strategy (IS) and Collaborative Design (CD). Four levels of Instructional Strategy were coded: (a) Media Delivery, (b) Behaviorist (traditional/teacher-centered), (c) Constructivist (student-centered), and (d) Independent Study. Collaborative Design was also coded into four levels: (a) no collaboration provided, (b) collaboration afforded (not required, voluntary use); (c) collaboration required, teacher moderated, and (d) collaboration required, not moderated (i.e., student facilitated). Table 17 displays the list of contrasts coded for each level of Instructional Strategy (IS) and Table 18 lists the contrasts coded for each level of Collaborative Design.

Comparison of Instructional Strategy Groups

The first group of results presented here are those for Instructional Strategy (IS). The results for the second group, Collaborative Design (CD), are presented later. All effect size statistics use the random effects model figures. Instructional Strategy (IS) was divided into four categories: The first category, called Media Delivery, included all studies in which no instructional contrast was made. These studies focused on the delivery media itself typically and attempted to hold all instructional aspects as equivalent as possible. The aggregate effect size for this group was g = .748 (k = 23, SE = .122). Group 2 included all studies in which behaviorist instructional strategies predominated. The aggregate effect size for this group was g = .812 (k = 27, SE = .150). Group 3 included all studies in which constructivist instructional strategies predominated. The aggregate effect size for this group was g = .698 (k = 17, SE = .182). Group 4

Table 17

Instructional Strategy (IS) Groups

Media Delivery	Behaviorist	Constructivist	Independent Study
Adcock et al 2006 1	Anderton 2005	Baturay & Bay 2010 1	Banks 2004 1
Adcock et al 2006 2	Boulter 2010 1	Baturay & Bay 2010 2	Banks 2004 2
Alavi, Marakas & Yoo 2002	Boulter 2010 2	Bernard & Lundgren 2001	Chang & Chang 2008
Benjamin et al 2008	Castaneda 2008 1	Bixler 2008	Clapano 2010
Caldwell 2006	Castaneda 2008 2	Cavus 2007 1	Collins 2000
Fox 2010 1	Castaneda 2008 3	Chen. C & Shaw 2006 1	Connolly et al 2007 1
Fox 2010 2	Castaneda 2008 4	Chen. C & Shaw 2006 2	Connolly et al 2007 2
Fox 2010 3	Castaneda 2008 5	Draper 2010 2	Cook et al 2007
Gupta 2006 2	Cavus 2007 2	Frey 2008	Draper 2010 1
Hairston 2007	Chen, B., Hirumi & Zhang 2007 1	Gulikers, Bastiaens & Martens 2005	Frith & Kee 2003
Hylton 2006	Chen, B., Hirumi & Zhang 2007 2	Krall et al 2009	Hansen 2000
Kanuka & Jugdev 2006	Chen, B., Hirumi & Zhang 2007 3	Lee 2010 2	Hansen 2008 1
Karatas & Simsek 2009	Gupta 2006 2	Mebane et al 2008	Hansen 2008 2
Karr et al 2003	Gupta 2006 3	Own 2006	Hansen 2008 3
Kemper et al 2006	Isenberg 2010	Peterson & Bond 2004	Jang et al 2005
Pacifici et al 2006	Jung et al 2002 1	Sendag & Odabasi 2009 1	LaRose, Gregg & Eastin 1998
Romanov & Nevgi 2006	Jung et al 2002 2	Yavuz 2007	Parsons 2006
Schroeder 2006 1	Lee 2010 1		Wallace et al 2006
Schroeder 2006 2	Ruksauk 2000		Williams 2005
Seabolt 2008	Shana 2009		
Sendag & Odabasi 2009 2	Stanley 2006 1		
Skylar 2004	Stanley 2006 2		
Westhuis, Oullette & Pfahler 2006	Tsai, Tseng & Hwang 2008 1		
	Tsai, Tseng & Hwang 2008 2		
	Wise et al 2004		
	Yang, Newby & Bill 2008 1		
	Yang, Newby & Bill 2008		

Table 18

Collaborative Design (CD) Groups

No Collaboration	Collaboration Afforded	Collaboration Moderated	Collaboration Facilitated
Adcock et al 2006 1, 2	Caldwell 2006	Alavi, Marakas & Yoo 2002	Bixler 2008
Anderton 2005	Cavus 2007 2	Baturay & Bay 2010 1, 2	Cavus 2007 1, 2, 3
Banks 2004 1, 2	Chang & Chang 2008	Bernard & Lundgren 2001	Jung et al 2002 2
Benjamin et al 2008	Chen. C & Shaw 2006 1	Castaneda 2008 1 - 5	Mebane et al 2008
Boulter 2010 1, 2	Collins 2000	Chen. C & Shaw 2006 2	Sendag & Odabasi 2009 1
Chen, B., Hirumi & Zhang 2007 1, 2, 3	Draper 2010 1	Draper 2010 2	Yavuz 2007
Clapano 2010	Hairston 2007	Frey 2008	
Connolly et al 2007 1, 2	Lee 2010 2	Frith & Kee 2003	
Cook et al 2007	Romanov & Nevgi 2006	Hylton 2006	
Fox 2010 1, 2, 3	Ruksauk 2000	Isenberg 2010	
Gulikers, Bastiaens & Martens 2005	Skylar 2004	Jung et al 2002 1	
Gupta 2006 1		Kanuka & Jugdev 2006	
Hansen 2000		Peterson & Bond 2004	
Hansen 2008 1, 2, 3		Sendag & Odabasi 2009 2	
Jang et al 2005		Shana 2009	
Karatas & Simsek 2009		Stanley 2006 1, 2	
Karr et al 2003		Tsai, Tseng & Hwang 2008 1, 2	
Kemper et al 2006		Westhuis, Oullette & Pfahler 2006	
Krall et al 2009		Yang, Newby & Bill 2008 1, 2	
LaRose, Gregg & Eastin 1998			
Lee 2010 1			
Own 2006			
Pacifici et al 2006			
Parsons 2006			
Schroeder 2006 1, 2			
Seabolt 2008			
Wallace et al 2006			
Williams 2005			
Wise et al 2004			

included all studies in which learners were left to their own devices. Instruction in this group did not include provision for student-student collaboration and very little contact with the instructor. Typically, students in these studies were given latitude to choose their own pace and sequencing of material. The aggregate effect size for this group was g = 0.848 (k = 19, SE = .212). These results are displayed in Table 19.

Table 19

Instructional Strategy (IS) Comparison

Instructional Strategy	k	g	SE
Media Delivery	23	0.748	0.122
Behaviorist	27	0.812	0.150
Constructivist	17	0.698	0.182
Independent Study	19	0.848	0.212
total	86		

These results would seem to indicate that the most effective instructional strategy when offering Web-based instruction is to let students teach themselves, i.e., Independent Study. The next most effective instructional strategy is behaviorist—which typically means high instructor guidance, one-way presentation of material and less student-centered practices. Because the media delivery category is not really an instructional strategy, it subsumes a variety of instructional practices and can be seen as a more or less base-line measure of Web-based instruction. The constructivist instructional strategy appears to be the least effective of the coded strategies.

Comparison of Collaborative Design Groups

The results for the second major coding group, Collaborative Design (CD), are presented next. CD category 1 included all studies in which no provision for collaboration was made in the instructional design. The aggregate effect size for this group was g = .808 (k = 40, SE = .110). Group 2 included all studies in which collaboration was afforded (made available) but not required as part of the instructional process. Participation and use of the collaborative tools were purely voluntary on the part of the learners. The aggregate random effects effect size for this group was g = .276 (k = 11, SE = .167). Group 3 included all studies in which collaboration was required and the instructor played a major role in the collaboration as a moderator. The aggregate effect size for this group was g = 1.049 (k = 27, SE = .159). Group 4 included all studies in which collaboration was required but the instructor played either no role or only a minor role in the collaboration. The learners were tasked with facilitating their own collaboration. The aggregate effect size for this group was g = 0.446 (k = 8, SE = .247). These results would seem to indicate that the most effective use of collaboration is when it is required and moderated by the instructor. Interestingly, studies examining instructional designs wherein no collaboration at all was afforded produced higher effects than did collaboration in which the students were responsible for making the collaboration work—and by quite a margin of difference. These results are displayed in Table 20.

Table 20

Collaborative Design (CD) Comparison

Collaborative Design	*k*	*g*	SE
No collaboration	40	0.808	0.110
Collaboration afforded	11	0.276	0.167
Collaboration moderated	27	1.049	0.159
Collaboration facilitated	8	0.446	0.247
total	86	--	--

Rank Order of Effectiveness for Instructional Practices Used in Distance Instruction

To answer the question: What is the most effective instructional practice used in Web-based, higher-education DE, the results presented will be rank-ordered according to effect size. Three such orderings will be made: a rank order of Instructional Strategies, a rank order of Collaborative Designs and a rank order of the two combined. Caution is urged when interpreting these rank orders as the effects of interactions and moderating factors have not been controlled for. Additionally, the combined rank ordering represents the ordering of two alternate groupings of the same sample, rather than an ordering of groups created from discrete samples.

Rank order of Instructional Strategies. If the largest effect size is indicative of the most effective instruction—that is, the IS that caused the largest growth in student achievement averaged over multiple studies, multiple samples and diverse circumstances (and no representation that such is the case is made here)—then the most effective IS used in Web-based higher education DE, as measured by the highest effect size, is Independent Study (*g* = .848). The next most effective—as determined by the next largest effect size—is Behaviorist (traditional or teacher-centered) instructional strategies (*g* = .812). The next most effective is Media Delivery (*g* = .748), followed by the least effective IS, Constructivist strategy (*g* = .698). It is important to note that this list is relative—that is, effectiveness is measured in relation to the other coded levels of instructional strategy as opposed to all possible instructional strategies used anywhere, at anytime. Even the strategy with the smallest effect size listed here—Constructivist (student-centered)—still appears to be very effective. The effect size for Constructivism (*g* = .698) represents an average growth of 2/3 of a standard deviation over the course many iterations of instruction.

Rank order of Collaborative Designs. In a similar fashion, the most effective Collaborative Design appears to be a design where collaboration is required and moderated by the instructor (*g* = 1.049). The cautions voiced concerning interpretation of Instructional Strategy voiced above apply equally to the ranking of Collaborative design. The next most effective design for collaboration appears, surprisingly, to be no collaboration at all (*g* = .808). Even more surprising is the difference in effect size between the two groups with the largest effect sizes and the two with the lowest effect sizes. The third most effective collaborative design is where collaboration is required, but it is left for the students to facilitate it and teacher input is minimal (*g* = .446). Least effective (comparatively) are designs where collaboration is afforded, but not required nor does the instructor participate in using it (*g* = .276).

Combined rank order. Finally, the two lists are combined to give an aggregate list of commonly used instructional activities ranked according to their relative effect sizes (Table 21). This combined ranking is more for the purposes of interesting comparison than as an analytical device to determine potential instructional effect. The most effective Web-based DE instructional practice used in higher education—based solely on the average effect size of the included studies in the present meta-analysis—is the use of moderated collaboration (g = 1.049). An effect size of 1.049 is very large, representing an average gain across all interventions measuring its effects, of more than one whole standard deviation. The next most effective methodology is Independent Study (g = .848). Note that, unlike some of the other coded categories, moderated collaboration and independent study are mutually exclusive by the coding protocols. The third most effective activity was the Behaviorist IS (g = .812). It, too, was mutually exclusive with Independent Study, but the majority of the Behaviorist group was formed by studies that used either moderated collaboration or no collaboration at all—the two next most effective methods on the list. Following the Behaviorist group in effectiveness was the group with no collaboration (g = .808), then media delivery (g = .748). Rounding out the bottom three are Constructivist (g = .698), facilitated collaboration (g = .446) and collaboration afforded (g = .276). Table 21 illustrates the rank order visually.

Summary of Results

This concludes the presentation of results according to the research plan. Prior to presenting the results of post-hoc testing and moderator searches, a brief summary of the above results is in order. The cautious answer to the first research question, "Which instructional interventions are most effective when used in a Web-based distance education setting—and under what circumstances?" is moderated collaboration, followed by independent study. Given the exploratory nature of this study, caution should be used before implementing these results. The somewhat more certain answer to the second research question, "Have Web-based distance education outcomes improved over time?" is, yes.

Table 21

Rank Order of Instructional Activities

Activity	*k*	Effect size
Collaboration Moderated	27	1.049
Independent Study	19	0.848
Behaviorist Strategies	27	0.812
No Collaboration	40	0.808
Media Delivery Only	23	0.748
Constructivist Strategies	17	0.698
Collaboration Facilitated (student led)	8	0.446
Collaboration Afforded	11	0.276

Part 3: Post-Hoc Tests and the Search for Moderators

Post-hoc Tests for Bias, Heterogeneity and Robustness

Post-hoc tests for heterogeneity. Three commonly used tests of homogeneity used in meta-analyses were performed on the entire included studies sample. These three tests are the Q test, the I^2 index and the τ^2 test. Each of these tests is briefly explained below.

The Q test, originally proposed by Cochran (1954) and later defined by Hedges and Olkin (1985, p. 123, Equation 25), has been the most commonly used measure of heterogeneity by meta-analysts. However, its power is directly dependent upon the number of studies or effect sizes included in the meta-analysis. In the present meta-analysis, the sample size of 59 studies falls in the middle range for sample sizes (see Huedo-Medina, et al. 2006), meaning that the Q statistic may be susceptible to Type I error. According to Huedo-Medina, et al. (2006), the Q statistic is problematic when used with the g effect size. They were unable to identify a suitable substitute, but implied that it is best to use multiple measures of heterogeneity when making statistical decisions. Typically, however, Q is used to test a meta-analysis sample to ascertain whether or not a fixed or a random effects model should be used. In the present case, it has already been assumed that the sample is highly heterogeneous and the random effects model will be employed.

The Q-test is limited to testing for the presence of heterogeneity, but interpreting the magnitude of the heterogeneity from the Q statistic is less-than straight-forward. Instead, a recently introduced statistic, I^2 (see Higgins & Thompson, 2002; Higgins et al., 2003) is used to easily determine the size of the heterogeneity. The I^2 index is calculated by taking the difference between the result of the Q test and its degrees of freedom (k -1), dividing by the Q value itself and multiplying by 100. It is interpreted as the percentage of the total variability in a set of effect sizes due to between-studies variability and thus, measures the amount of true heterogeneity.

A final measure of heterogeneity commonly used in meta-analyses is the τ^2 test which estimates the between-studies variance. In a meta-analysis using a random effects model, the between-studies variance is a reflection of how much the true population effect sizes estimated in each of the single effect sizes in a meta-analysis differ from each other. One drawback to τ^2 is that it cannot be generalized between meta-analyses that use different effect size measures.

Homogeneity of the meta-analysis sample. Results of these three tests for homogeneity for the entire included studies sample are shown in Table 22. As anticipated, the included study sample is highly heterogeneous ($Q_{(83)} = 903.678$, $p < .001$). According to the I^2 index, 90.815% of the total heterogeneity in the sample is due to between-studies differences, a result to be anticipated when combining studies from diverse sources, times, fields of study and involving highly diverse populations. Though somewhat redundant in the present case, the τ^2 test was also conducted. Results indicate considerable between-studies heterogeneity ($\tau^2 = 0.430$, SE = 0.130). These results confirm that all inferences made regarding the outcomes of meta-analytic statistical calculations should use figures from the random effects model only. The results also suggest the presence of many small sample size effect sizes with relatively large individual effect sizes. This is not necessarily a negative, as Sterne and Egger (2001) note. They observe that tests such as these do not assign causality, only relationship. That is, it is entirely plausible that the effect size in smaller studies may be larger because the effect is, in fact larger. While it is possible that the effect size in smaller studies may be larger because of publication bias, it is equally likely that the presence of large effect sizes in smaller studies is because those studies use different populations and/or different protocols or exert better control over confounds than possible in larger studies. In fact, Song et al. (2002), in a study of 28 meta-analyses, found that smaller sample size studies had greater accuracy than large studies. Thus, the presence of many small studies with large sample sizes is not necessarily due to publication bias. To ascertain the case in the current study, additional post-hoc tests for publication and selection bias was run. First, however, one final test for heterogeneity was run to see if it is possible in the present case to reduce the heterogeneity of the sample.

Table 22

Tests for Homogeneity for the Entire Sample of Included Studies

Sample Heterogeneity				Tau-squared			
Q	df (Q)	p	I^2	τ^2	SE	Variance	τ
903.678	83	0.000*	90.815	0.430	0.130	0.017	0.656

* significant at α = .001

One-study removed check for outliers. One common post-hoc test performed when a meta-analytic sample is heterogeneous is the one-study removed test. Because extreme heterogeneity can sometimes lead to erroneous interpretation of the results, the one-study removed test is a strategy used to identify the presence of one or more extreme outlier effect sizes that may be skewing the results one way or another. Comprehensive Meta-analysis automatically performs this test reiteratively. That is, it consecutively removes one effect size at a time, recalculates the aggregate statistics and then compares all the individual results to determine which single study being removed has the greatest effect on the outcomes. The software then reports that as the one study removed results. The results of the one-study removed check for this study are shown below in Table 23.

Table 23

One-Study Removed Test for the Impact of Heterogeneity on the Main Effects

	Effect size and 95% confidence interval					Test of null (2-Tail)	
k	Point estimate	SE	Variance	Lower limit	Upper limit	Z	*p*
84	0.768	0.078	0.006	0.615	0.921	9.864	0.000*

* significant at $p < .001$

The results of the one study removed test indicate that no single effect size has an undue effect on the entire sample outcomes. With one study removed, the effect sizes are still significant and remain close to the values yielded by the entire study. A comparison of the one-study removed values with the original values (see Table 11), shows that the overall effect size is only slightly reduced by removing the most extreme outlier value from the sample (g = .777, SE = .078, variance = .006 for the original value versus g = .768, SE = .078, variance = .006 for the one-study removed value). Thus, heterogeneity is widely dispersed among all the studies as opposed to being concentrated in a few highly extreme outlier studies.

Having determined that the study sample is, in fact highly heterogeneous—as expected—and that it is composed of a number of smaller sample size effect sizes of relatively large magnitude, it was necessary to determine whether or not the sample is compromised by publication bias or selection bias.

Post-hoc tests for publication/selection bias. Two dangers inherent in the meta-analytic process are selection bias and publication bias. Though the two are functionally different, they have the same operational effect on a meta-analysis: they “produce” missing studies. Since a meta-analysis is designed to aggregate all possible relevant studies that fit the inclusion criteria, it becomes problematic whether a truly exhaustive meta-analysis is possible. The problem is summed-up in the classic “file drawer” illustration used by Robert Rosenthal (1979) to explain the improbability of a truly exhaustive meta-analysis: Somewhere in a file drawer there exists a long forgotten study that was either inconclusive or that found no significant difference that was either never submitted for publication or was rejected. Not knowing that such a study exists and there being no record of it in any accessible database, it is unlikely to ever be found. Yet, its very inconclusiveness is an important contribution to the overall results of an exhaustive meta-analysis. The question is: how many such studies exist? If there were enough such studies, their aggregate effect might be enough to nullify any aggregate effect size among the studies that are located.

Rosenthal suggested that rather than be concerned with trying to achieve the impossible by locating all such studies, it would be better to calculate the number of such studies that would be required to reduce any given main effect to zero. If the number of such studies was large, then it could be safely assumed that the study sample used in the meta-analysis was adequate to produce a reasonably accurate estimate of the actual effect. If the number of such studies required to reduce the effect to zero was small, then the meta-analytic study sample was inadequate. One by-product of this insight was the check it also provided against selection bias because it doesn't really matter why a study is missing (publication bias, selection bias or search inadequacy)—calculating the effect of missing studies on the outcome of the sample that is present provides a check on all of them. There are a number of post-hoc tests commonly used by meta-analysts to check for the presence and effect of missing studies on the outcomes of the meta-analysis. Unfortunately, none of them are reliable when large between-studies variability exists in the sample, as is the present case. One of the most popular tests, possibly because of its visual nature, is the funnel plot.

Funnel plot. A funnel plot is a graph plotting study size (using either the standard error or precision—the reciprocal of the standard error) on the vertical axis versus effect size on the horizontal axis. In this type of plot, the larger a study is (i.e., the larger the Standard Error), the higher along the vertical axis it appears and the larger its effect size, the farther to the right along the horizontal axis it is located. Thus, small studies with large effect sizes tend to cluster in the lower right-hand corner. If no studies are missing, it is expected that the studies would be distributed symmetrically around the mean effect size. An asymmetrical distribution results whenever this assumption of centrality is violated—usually because of missing studies. It doesn't matter whether the studies are missing due to publication bias or selection bias or the failure to conduct a truly exhaustive search: a missing study is a missing study and a funnel plot can identify the presence—do to speak—of missing studies (Egger et al., 1997).

Figure 2 presents the funnel plot for the current study. The funnel plot appears to be asymmetrical, indicating the presence of missing studies on the left-hand side, but not necessarily the presence of publication bias, since there are few small studies with large effect sizes (i.e., studies falling in the lower right-hand corner). It is difficult to ascertain whether the plot for this study is truly asymmetrical without testing it statistically. Two commonly used statistical tests are used to determine whether true asymmetry exists as opposed to simply appearing to be asymmetrical. In a sense, these measures test for the presence of a statistically significant asymmetry much in the way that other inferential statistics measure the significance of the differences between group means.

Begg and Mazumdar Rank Correlation. The first statistical test reported here is the Begg and Mazumdar Rank Correlation (1994) test. It computes Kendall's *tau*-b (rank order correlation) between the effect and the standard error (again, a function of sample size). The Cochrane Commission cautions that this test has low power and frequently does not detect bias. A significant correlation can suggest that bias exists but cannot tell anything about it. Essentially, a significant correlation is confirmation that asymmetry exists in the sample, but does not indicate the source of that asymmetry. In the case of the current study, the Begg and Mazumdar test detected statistical asymmetry in the included studies samples ($p < .05$; 1-tailed), but only if an alpha of .05 is assumed (see Table 24).

The distribution is asymmetrical, but not extremely so. About half the studies fall near the mean effect size with more falling to the right of the mean than the left. Two extreme outliers, one a small size study with a small effect size and the other a medium size study with a large effect contribute to the visual impression of asymmetry. The

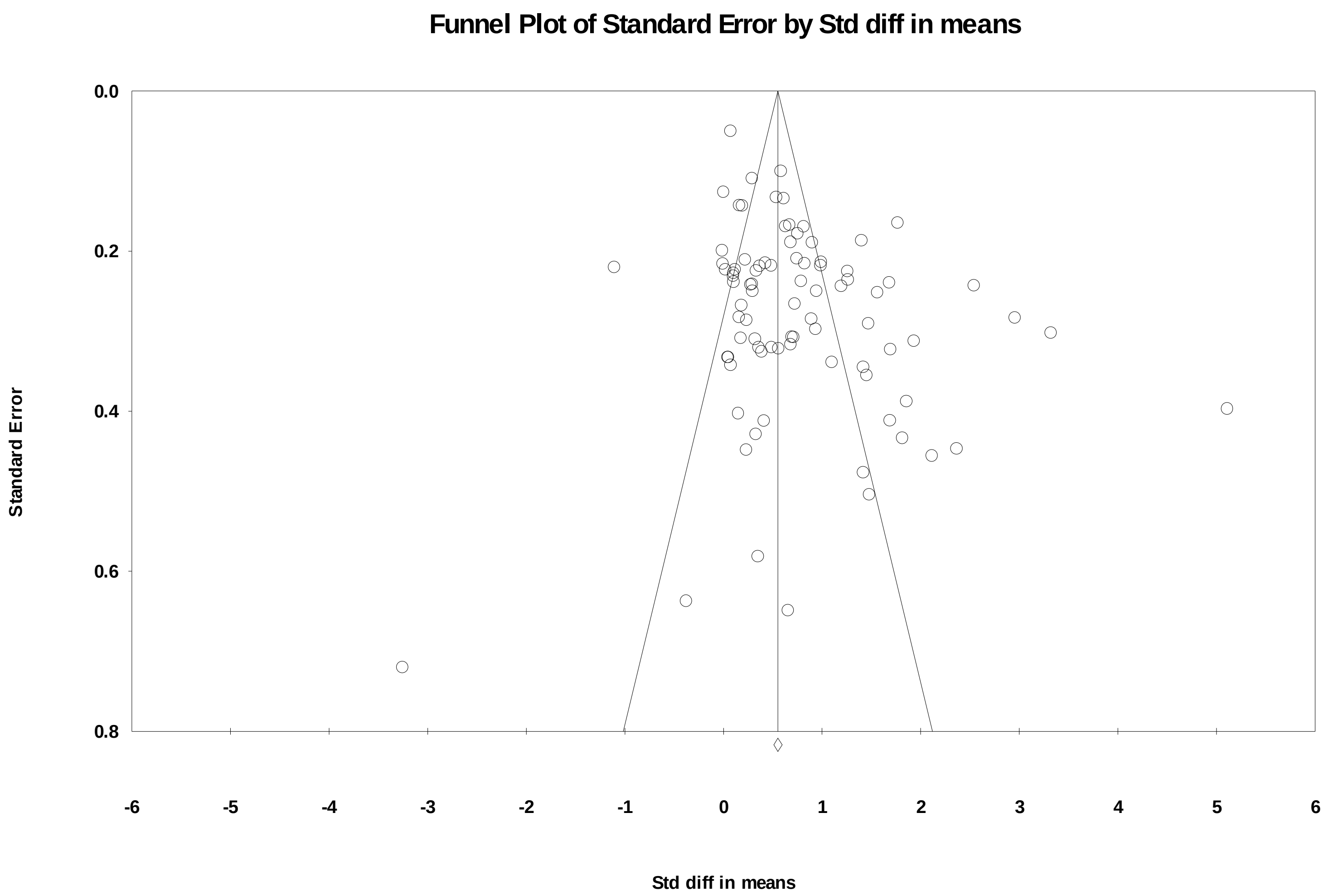

Figure 2. Funnel plot of the included studies. Source: Comprehensive Meta-Analysis.

Table 24

Begg and Mazumdar Rank Correlation Test for Asymmetry

Kendall's S statistic (P-Q)	519.0000
Kendall's tau without continuity correction	
tau	0.14200
z-value for tau	1.93579
p-value (1-tailed)	0.02645
p-value (2-tailed)	0.05289
Kendall's tau with continuity correction	
tau	0.14172
z-value for tau	1.93206
p-value (1-tailed)	0.02668
p-value (2-tailed)	0.05335

suspicion of bias suggested by the funnel plot is weakly confirmed by this test. However, it is important to remember that one of the weaknesses of these post hoc tests is their unreliability when dealing with highly heterogeneous samples like the present one, so the test is hardly conclusive. In the case of inconclusive results such as these, it is best to use multiple measures as a check. Accordingly, a further test for bias was conducted.

Egger's test of the intercept. A somewhat more powerful test than the Begg and Mazumdar test is Egger's Test of the Intercept (Egger et al., 1997). Egger simply emulates the Begg and Mazumdar test but uses precision (the inverse of the standard error, i.e., 1/SE) rather than the standard error itself. In the present case, Egger's test also finds that the sample is biased ($p < .001$), using a more rigorous standard ($\alpha = .001$), see Table 25. Thus, it can be fairly confidently assumed that some sort of bias in the sample exists, but the lack of a large number of small sized studies with large effect sizes suggests that the bias is due to some other factor than publication bias. Fortunately, it doesn't matter what the source of the bias is, because there are at least three methods available to meta-analysts to correct for sample bias problems.

Table 25

Egger's Regression Intercept Test for Sample Bias

Egger's regression intercept	
Intercept	2.64703
Standard Error	0.67378
95% lower limit (2-tailed)	1.30715
95% lower limit (2-tailed)	3.98691
t-value	3.92863
df	84.0000
p-value (1-tailed)	0.00009
p-value (2-tailed)	0.00017

Trim and Fill. The classic correction for sample bias in a meta-analysis is Duval and Tweedie's Trim and Fill method (2000). This method begins by iteratively removing asymmetric studies from the right-hand side of the funnel plot calculating the main effect until only unbiased effects remain. The procedure then replaces the removed effects on *both* sides of the mean effect size to create an imputed symmetry. Table 26 shows that the trim and fill procedure determined that—theoretically—24 studies are missing from the current study sample. While that is entirely possible, the low rate of return (a fraction of a percent) from the search for studies makes it unlikely that 24 studies that meet the criteria for inclusion in the present study actually exist—in file drawers or elsewhere. Since it is unlikely that 24 studies can be located, the question is whether their absence makes any difference and, if so what to do about it. That is where Rosenthal's file drawer solution comes into play.

Table 26

Duval and Tweedie's Trim and Fill

Studies Trimmed	Random Effects Model Point Estimate	Lower Limit	Upper Limit	*Q*
	0.77659	0.62352	0.92964	911.22298
24	0.34891	0.17527	0.52256	1790.01666

Top row: observed values
Bottom row: Adjusted values

Fail-safe N. Harris Cooper (1979) proposed the term "fail-safe N" as the name for the statistic suggested by Rosenthal to address the file-drawer problem. There are currently two ways of computing the fail-safe N. The first, known as the classic method, is to compute an effect size for each study/contrast, combine the effect sizes and compute the *p*-value for the combined effect. A second method was suggested by Orwin (1983). Unlike the original method proposed by Rosenthal, Orwin's method assumes that the missing studies may include studies with effect sizes in the reverse direction, not merely null effects. It allows for the user to set two parameters: the mean value of the effect sizes of the missing studies (zero is the default value in CMA) and the target value of the effect size for the combined existing sample plus the missing studies. The larger the fail-safe N, the more representative the meta-analysis sample is and the more robust the results. Table 27 presents the results of the Fail Safe N for the present study. The classic fail-safe N for this study sample is 6634 studies and Orwin's fail safe N, with a criterion of 0 mean effect size in the missing studies and a target threshold of .001 effect size would require 7481 such studies. By contrast, Bernard et al. (2009) required only 44 studies using the classic fail safe N (also computed by CMA) to reach null value. None of the other meta-analyses mentioned in the introduction to this study—including Bernard et al. (2004)—reported a fail-safe N. In a meta-analysis of 91 clinical trials, Doughtery and Done (2009) reported an Orwin's fail-safe N of 41 using a cut-off criterion of .20. By contrast, using the same parameters, the present study would require 152 studies with a zero effect size to reduce the main effect to 0.20. In another way of looking at the impact of bias, if all 24 studies that the Duval and Tweedie's trim and fill test identified as being missing were added to the included studies and each had an effect of -1.0, the main effect for the combined studies would still not drop to 0.20. In other words, the existing included study sample provides an extremely sound representative sample of the extant studies meeting the criteria for inclusion.

Summary and conclusion for post-hoc testing. Based on the strength of the fail-safe N tests, as well as the relative weakness of the findings of asymmetry that suggest a bias in the study sample, it seems unlikely that the sample is biased in any way. It is highly heterogeneous, which makes absolute determination of bias uncertain in either direction, but the strength of the fail-safe N suggests two things about the sample:
First, the sample though heterogeneous, is more than adequately representative of the extant studies and a truly exhaustive sample would be unlikely to differ much in its

outcomes from the present study. This is important for generalization of these findings beyond the present sample.

Table 27

Classic and Orwin's Fail Safe Ns

Classic fail-safe N	
Z-value of observed studies	27.32803
p-value for observed studies	0.00000
Alpha	0.05
tails	2
Z for alpha	1.95996
Number of observed studies	86
Number of missing values to bring p-value to alpha	6634
Orwin's fail-safe N	
Standard difference in observed studies	0.55310
Criterion for a "trivial" std diff in means	0.00
Mean std diff in means in missing studies	0.00
Number of missing studies needed to bring std difference in means under 0.001	7481.000

Second, the apparent lack of bias coupled with the relative absence of small study sizes with large effect sizes suggests that what asymmetry there may actually be is more likely attributable to actual strong positive treatment effects or other systematic errors at the study level, than to missing studies or systematic error at the meta-analytic level. That is, the strength of the main effect found by this meta-analysis is likely to be a reasonably accurate estimate of the actual effect of the treatments measured by the constituent research studies.

Identifying Moderators

As part of the coding scheme for this meta-analysis, four potential moderators were coded for each study: One moderator for the IS strategy category Media Delivery, two moderators for the IS category Behaviorist, and one moderator for the IS category Constructivist. In addition, where appropriate, the moderator categories for Behaviorist were coded for independent study as well. The results of moderator testing for each category of IS follows.

Moderators of Media Delivery. Of the three categories of moderator values coded for media delivery, only one was found to be significant: Multimedia (Q_b = 41.269; k = 5, p = 0.000). In addition, an instructor role moderator, Formative Feedback was also a significant moderator (Q_b = 14.463; k = 4, p = 0.002). Table 28 presents all the relevant information. Interestingly, interactive multimedia was not a significant moderator and no text-only comparison studies were coded.

Moderators of Constructivist strategy. Four levels of a single moderator, Instructional Activity, were coded for the Constructivist instructional strategy. One of the four levels, Problem-Based Learning, was identified as a significant moderator of the effect size for this category (Q_b = 46.007; k = 8, p = 0.000).

Additional moderators. No moderator values were coded for Collaborative Design, as the four levels of the category were considered to be sufficient to identify the specific instructional contrasts associated with this coding variable.

Attempts to drill-down beyond the main moderator categories for the behaviorist and constructivist instructional strategies proved to be impossible; there were simply not enough independent effect sizes to form groups large enough or diverse enough for the

Table 28

Moderators of Media Delivery

		Effect Size			95% confidence interval		Heterogeneity		
Group	*k*	*g*	SE	Variance	Lower limit	Upper limit	Q_b	Q_w	*p*
Multimedia	5	1.027	0.399	0.159	0.245	1.809		41.269	0.000
Interactive	6	0.771	0.114	0.013	0.549	0.994		8.259	0.143
Formative Feedback	4	0.870	0.282	0.079	0.317	1.422		14.463	0.002
Deliverer of Content*	5	0.302	0.143	0.020	0.022	0.582		10.845	0.028
Total Between	20						8.536		0.036

* Equivalent to no active role or no role described, this is the default category when no other moderator category could be coded for media delivery.

comparisons to have any meaning. For instance, only 19 effect sizes derived from 14 studies provided enough information to code for Instructional Activity (IA) categories 1–4. In contrast, 67 effect sizes from 45 different studies were coded as IA category 5 (other) or IA category 6 (not identified), with the majority of those being coded as “not identified.” In addition, an attempt was made to identify interaction effects between IR and but once again the lack of codeable data prevented the formation of viable moderator sub-groups.

Summary of Findings

To summarize, this study found the following:

1. Main Effect. The main effect of Web-based DE instruction since 1998 on student outcomes was g = .777 (k = 59, SE = .078). This is the actual effect of instruction on student outcomes, not the difference in effect between f2f and DE as has most often been reported in past meta-analyses.
2. Improvement over time. The effectiveness of Web-based DE appears to have increased over the past 13 years from a mean effect size of $g = .606$ ($k = 19$, SE = .153) for studies published prior to 2006 to a mean effect size of $g = .830$ ($k = 18$, SE = .172) for studies published in 2009 and after.
3. Instructional Strategy. The most effective instructional strategy for higher education Web-based DE, based on effect size only, appears to be Independent study ($g = .848$, $k = 19$, SE = .212), followed, in order, by Behaviorist ($g = .812$, $k = 27$, SE = .182), Media Delivery ($g = .748$, $k = 23$, SE = .122) and Constructivist ($g = .698$, $k = 17$, SE = .182).

4. Collaborative Design. The most effective collaboration design for higher education Web-based DE appears to be collaboration moderated by the instructor (g = 1.049, k = 27, SE = .159), followed, in order, by no collaboration provided (g = .808, k = 40, SE = .110), mandatory student facilitated collaboration (g = .446, k = 8, SE = .247), and voluntary use of collaboration (afforded) (g = .276, k = 11, SE = .167).
5. Instructor Role. The provision of formative feedback by the instructor was the most significant modifier of the effect of both media delivery studies (Q_b = 14.463; k = 4, p = 0.002) and behaviorist instructional strategies (Q_b = 30.419; k = 10, p = 0.000).
6. Instructional Activity. The use of multimedia delivery techniques was a significant modifier of media delivery study outcomes (Q_b = 41.269; k = 5, p = 0.000), as was simulations for Behaviorist instructional strategies (Q_b = 15.837; k = 5, p = 0.003) and Problem-Based Learning for Constructivist instructional strategies (Q_b = 46.007; k = 8, p = 0.000).
7. Sample quality. The data sample upon which these findings are based is highly heterogeneous ($Q_{(83)}$ = 903.678, p < .001; I^2 = 90.815) but representative and very robust according to Orwin's Fail-safe N (7481).

CHAPTER 5

DISCUSSION AND CONCLUSIONS

Chapter Five includes the discussion of the findings, post hoc explanatory statistical analysis of those findings and conclusions based on them, followed by the implications of those conclusions. It includes most of the information listed under the heading of "Discussion" in MARS. The chapter also includes some non-statistical observations regarding the material and processes encountered in the course of the research for the study and ends with some suggested guidelines for future research.

This study seeks to answer two research questions in regard to the use of Web-based distance education. The two questions are re-stated here:

Research Question 1: Which instructional interventions are most effective when used in a Web-based distance education setting—and under what circumstances?

Research Question 2: Have Web-based distance education outcomes improved over time?

Discussion

Main Effect

As reported previously, the main effect for the effect of Web-based distance instruction on student outcomes is g = .777 (k = 59, SE = .078). It is based on contrasts reported by high-quality studies that control for media delivery and study quality as inclusion parameters and it holds group equivalence as a necessary pre-requisite for calculating the effect of instruction. Post-hoc tests suggest that though the included studies are highly heterogeneous, bias in the inclusion of studies is minimal with the result that the reported effect size is highly robust. In other words, the effect size reported here is about as reliable and accurate a figure as is likely to be derived from the studies extant on December 2010. Given that, what does that effect size actually mean?

What does the main effect size mean? An effect size of .777 signifies that the mean of the treatment group is almost one standard deviation higher than the mean of the control group. This difference is illustrated graphically in Figure 3. The typical way of interpreting effect sizes is to use Cohen's (1988) benchmarks. These benchmarks for effect sizes have, unfortunately, become uncritically accepted as the de facto method of labeling effect sizes, but Cohen never intended for them to be used in that fashion. As usually presented, Cohen's benchmarks for effect sizes are: d = .20 or r = .10 is a *small* effect size; d= .50 or r = .30 is a *medium* effect size and d = .80 or r = .50 a large effect size. He intended these to apply to the behavioral sciences as a whole, but cautioned that specific fields within the behavioral sciences could have distinctly larger or smaller effects sizes as a norm. Using Cohen's benchmark as a guide, the main effect for Web-based instruction reported in this dissertation can be considered a "large" effect.

Cohen provided examples to explain what he interpreted as a small, medium and large effect. For instance, he likened a small effect to be the difference in height between 15-year-old and 16-year-old girls in the US. In a similar way, a medium effect size is one "large enough to be visible to the naked eye (Cohen, 1969, p. 23)," for instance, the difference between the heights of 14-year-old and 18-year-old girls. He described large effect sizes as "grossly perceptible," similar to the difference between the heights of 13-year-old and 18-yearold girls. In an appropriate example, he suggested that an effect size of .8 would approximate the difference in performance on an IQ test between holders of the Ph.D. degree and "typical college freshmen (Cohen, 1969, p23)."

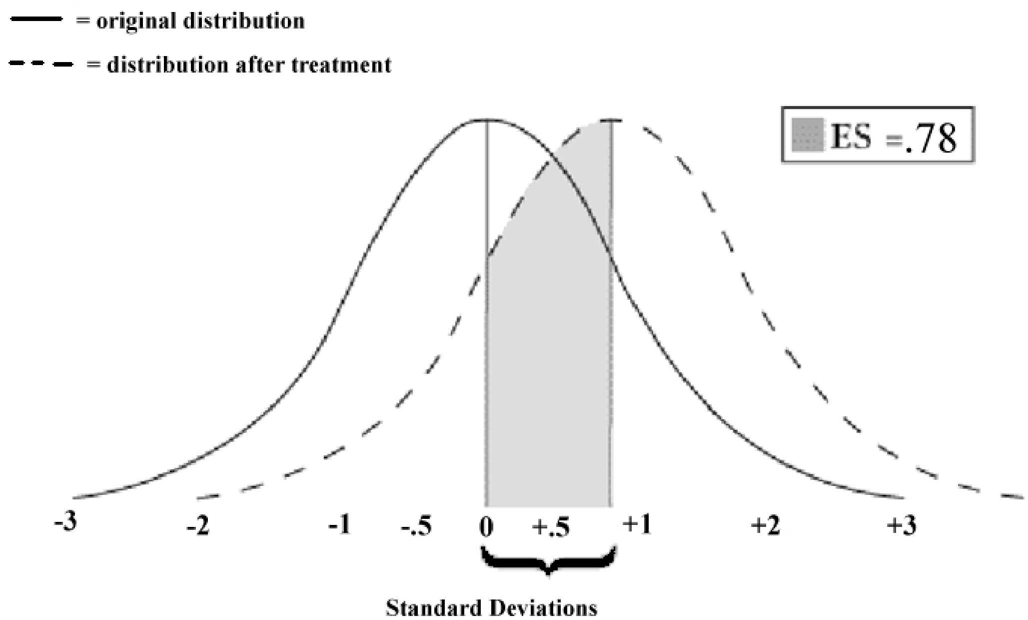

Figure 3. Graphic depiction of main effect size for Web-based distance education. Adapted from Marzano (1998).

There are a number of other ways to interpret an effect of g = .777. Thought of as a Z-score, an effect size of .777 means that the score of an average person in the treatment group is almost .8 of standard deviation higher than that of the average person in the control group. Another way of interpreting the score is that about 78% of the individuals in the control group would score below the mean in the treatment group or that the 6th highest score in the control group would only be equal to the mean in the treatment group.

Two additional methods of interpreting the main effect size are the Binomial Effect Size Display or BESD (Rosenthal & Rubin, 1982) and the Common Language Effect Size (CLES) suggested by McGraw and Wong (1992). The BESD is a metric designed to quantize the difference in the percentage of successful treatments (usually defined as greater or lesser than the median score of the combined groups) between the control and treatment groups. In other words, the BESD for the main effect of Web-based instruction for this meta-analysis (BESD = .36) means that 36 % more individuals in the treatment group scored greater than the median score than in the control group. The CLES is designed to make sense to non-statisticians and is the likelihood of the score of a randomly selected individual from the treatment group being higher than a randomly selected individual from the control group. The CLES for the current main effect is .70 which means 70 out of 100 times, the score of the person randomly selected from the treatment group in the this study would be larger than the score of a random selection from the control group.

Contrasts between this study and previous studies. As mentioned briefly before, there is a distinct difference between the main effect reported in this study and those typically reported in earlier meta-analyses. First, the effect size reported here (g = .777) is the actual mean effect size for the difference between the outcomes of treatment versus the contrast group outcomes, either a baseline figure for a within-groups contrast (pre-post test) or the outcomes from a control/contrast group in a between-groups contrast (equivalent groups). Thus, this effect size measures the absolute effect of the treatment on the treatment group rather than the relative difference between two differing treatment groups as is the case when comparing DE to f2f. Because of this, the main effect figure reported here is expected to be larger than that reported by earlier meta-analyses. The result is as expected: The main effect for this study is almost twice as large as that for Bernard et al. (2009), the most comparable study.

A second contrast between this meta-analysis and previous meta-analyses is that it is the only meta-analysis to completely control for media delivery by comparing only studies using the same type of media delivery—with no mixing of delivery types. Even Bernard et al. (2009) who compared only DE studies to DE studies still confounded media delivery by including ITV and other distance delivery methods with Web-based delivery. The current study is the only study to restrict inclusion to Web-based delivery only. Every study included in this meta-analysis involved instruction delivered via the World Wide Web. Thus, media delivery—though it has been shown to have either no or negligible impact on outcomes—is not a potential systemic contributor to variation between studies. This means that there can be no direct comparison between the outcome effect sizes of this meta-analysis and that of any other meta-analysis thus far located. The closest comparison may be with the results of a moderator of computer use reported by this author in an earlier meta-analysis (Roberts, 2002). In a meta-analysis comparing different ways computers are used (the model upon which this current meta-analysis is based), the use of the computer for distance education showed the largest effect size (d = 1.56, k = 2, χ^2= 20.69). Because this effect size was based on only two studies, that effect size needs to be treated with caution, but serves for at least some point of comparison (Roberts, 2002).

Improvement over Time

This dissertation found that the effectiveness of Web-based DE has apparently improved over time. There are likely to be a number of reasons for this, not the least of which is that the design of this study makes it difficult to precisely compare the current results with those of past research syntheses. Quite apart, however, from comparisons to previous meta-analytic findings is the relative increase depicted solely within the confines of this study. That is, the finding of increased effectiveness is based on the difference between groups of studies in this meta-analysis—all included according to the same criteria, all using the same methods of aggregation and the same effect size calculation. It is clear that there is a trend for larger effect sizes over time in the present study, regardless of what other analyses may or may not have found. Explaining this increase over time on purely systemic methodological grounds is difficult to do when there are no methodological differences between the groups—apart, possibly, from the selection of dates used to separate the chronological groups. More likely, the differences can be explained by multiple factors that affect the between studies differences.

Among the factors most likely to contribute to the increase in mean effect size of Web-based instruction over time are advancements in the technology used to deliver Web-based instruction, an increase in experience and training on the part of instructors and a commensurate increase in experience, familiarity and comfort with Web-based delivery of instruction on the part of learners. In the absence of additional research examining the presence of and potential effect of improvements in technology, instructional techniques and student proficiency with the methods and tools of distance instruction, there is little to be gained from further speculation into the potential causes for the observed increase in effect size over time. This is definitely an area of research needing further attention.

Instructional Strategy

The central concern of this study was identifying instructional practices that appear to be most effective when used to deliver Web-based instruction. Prior to discussing the results, it must be emphasized that this is an exploratory study, intended primarily to identify the profitable areas for more in-depth research. The implications for practice need to be carefully weighed before potential implementation.

Given the attention accorded to constructivist and collaborative instructional practices the past few years—especially in respect to the use of digital computing technologies (e.g., Tam, 2009; Wiburg, 2009)—the results of this study were surprising. Constructivist instructional strategies produced lower effect sizes than the other three categories of instructional strategy. Perhaps most surprising—and disturbing—was that independent study appeared to be the most effective method of instruction. The idea that leaving students to their own devices results in greater achievement than intervention by trained instructors is both counter-intuitive and troubling in its implications. It is less difficult to believe that the "tried and true" methods of Behaviorist teaching strategies might be more effective than relatively new Constructivist methods, as seems indicated by the results of this meta-analysis. The hierarchy of effect sizes observed in the results of this meta-analysis is: independent study (highest effect size), Behaviorism, media delivery followed by Constructivism (lowest effect size). Each of these categories is discussed below.

Independent study. The most puzzling outcome of this study is the implication that independent study designs might lead to higher student achievement than do strategies that incorporate more interaction between humans. Using the basic interaction types described by Bernard et al. (2009), studies categorized as independent study in this study were solely or predominately student-content interactions in nature, with little or no student-student or student-instructor interaction. Bernard et al. (2009) found that both student-student and student-content interaction yielded significantly larger observed mean effect sizes than student-teacher interaction and that there was no significant difference between student-student and student-content interaction. Moreover, they also found that "only strengthening SC [student-content] was related to increasing effect size (p. 1265)." They concluded that course design features that help students engage in content make a "substantial difference" in achievement (p. 1265). In addition, they found that "the relationship between the strength of ITs [Interaction Treatments] and achievement held for asynchronous DE courses but did not hold for 'not asynchronous' courses (p. 1265)."

The results of this meta-analysis would seem to confirm the suspicions of Bernard et al. 2009 in that studies categorized as independent study can also be thought of as largely—if not entirely—asynchronous. That is, the learner and the instructor are not physically present in the same place nor does interaction between the two normally take place simultaneously. Thus, courses that are designed—intentionally or otherwise—to be independent studies may tend to focus all activities on student-content interaction to the exclusion of other types of interaction and in doing so, may possibly strengthen the instructional efficacy of the design.

Another possibility for these results lies in the somewhat controversial idea of andragogy. One of the central ideas in andragogy, which is understood to refer to learning or instructional principles that are particular to adults and are either different from or a continuation of pedagogical principles directed primarily at children. It is not the intent of the current study to enter into the debate about whether and to what extent andragogy exists, but to suggest that certain elements attributed to it may explain the surprising results observed for independent study. Central to the concept of andragogy is the idea of self-directed learning (SDL); that is, adult learners are in charge of their own learning and learn best when they control that learning. At odds with this independence is the notion that SDL should be highly collaborational—which is not what one envisions when thinking of independent study. There are, in fact, two schools of thought in regards to SDL, one oriented toward the individual (e.g., Braman, 1998; Long, 1994; Merriam & Caffarella, 1999) and one oriented toward collaboration (Maehl, 2000; O'Donnell, 1999; Rowland & Volet, 1996). The thrust of most of the collaborative views seems to be that self-direction is learned or obtained through collaborative interaction. Thus, one might look at SDL as more a matter of maturation, than one of preference: the more mature a learner, the more likely that SDL is an effective and appropriate learning strategy. The implication, of course, is that independent study should be reserved for the most mature, self-efficacious learners and, if such is the case, then high achievement (i.e. large effect sizes) should be expected. Unfortunately, the majority of subjects in the studies coded as independent study were undergraduates, so no conclusions regarding the idea of maturity could be drawn from the demographic statistics. It is possible, however, that the maturational difference is between children and adults—again reflecting the idea of andragogical explanations.

A similar, but not altogether identical, concept to self-directed learning and one primarily identified with European countries is *autodidacticism* meaning “self-teaching.” It is primarily identified with highly successful individuals who teach themselves and go on to “prove” the efficacy of their self-teaching through their outstanding accomplishments. Implicit in the idea, of course, is the existence of many autodidacts who were less outstanding in their accomplishments, but no less successful in teaching themselves. It is possible that some factor similar to autodidactivism is operating in the Web-based courses herein coded as “independent study.” The results observed in the present study suggest that, at least among adult learners, affording control over their own learning is not only effective, but more effective than other instructional strategies.

The current results suggest that more directed research into the efficacy of andragogical orientations to instruction for Web-based instruction aimed at adult learner might be profitable. In any case, it appears that giving control to students in Web-based instruction may lead to higher achievement.

Behaviorism versus Constructivism. The largest source of debate highlighted by the results of the present study is the apparent effectiveness of Behaviorist strategies over Constructivist strategies. There are three areas of discussion that seem pertinent to the present case: (1) there is very little empirical research directly comparing Behaviorist instructional outcomes to Constructivist outcomes, (2) fidelity of implementation of Constructivist instructional techniques are still problematic, and (3) evidence suggests that Constructivist instructional techniques may not produce outcomes that are aligned with the quantitative measures employed in many of the studies reviewed.

The first consideration in comparing the results of the current study with respect to Behaviorist strategies and Constructivist strategies is the relative lack of empirical research directly comparing the two. In a recent meta-analysis, Rosen and Salomon (2007) compared constructivist and “traditional” instruction and found a main effect of .460 for Constructivist practices compared to traditional instructional practices, but the differences between the two disappeared when only traditionally-appropriate outcome measures were used. In contrast, when Constructivist-appropriate outcome measures were used, the effect size in favor of Constructivist instruction rose to .902. It appears that Constructivist instruction is highly sensitive to differences in outcomes measures that may partially explain the results observed in this meta-analysis.

In a study of 10th grade students in Turkey, Akkuş, Kadayifçi, Atasoy and Geban (2003) compared constructivist instruction to traditional instruction in science and found that post-test scores for the Constructivist group were significantly higher than that of the score in the traditional group. Unfortunately, the study used two existing classrooms without group equivalence being established and did not indicate how the treatment classroom was determined. Moreover, the treatment classroom focused a good portion of its instruction on identifying and eliminating false scientific pre-conceptions—a learning strategy that is commonly taught as pedagogical-content knowledge in the area of science today and is not necessarily a Constructivist instructional strategy.

Several studies have pointed out that though Constructivist instruction seems to lead to improved recall and greater understanding, it requires more time to do so (Lord, 1997; Tynjälä, 1999; Yuen & Hau, 2006). In other words, Constructivist instruction focuses on smaller amounts of in-depth learning as opposed to traditional (Behaviorist) methods which appear to be more effective at covering a broad spectrum of topics in lesser depth. Thus, the operational instructional principal is time-on-task rather than differences in instructional approach. Thus, it might be said that Constructivist methods focus on less material covered in-depth and Behaviorist methods concentrate on breadth of material at a lesser depth of understanding. The consequences are obvious. If an outcome measure is designed to measure the type of knowledge typically taught by Behaviorist methods, then Constructivist methods will not measure-up as well—and vice-versa. In the studies contained in the present study, all the classes were DE adaptations of existing f2f classes originally taught by traditional, behaviorist methods and the outcome measures used for those versions of the class were also used to measure the outcomes of constructivist instruction. The lower effect size (still a large effect size) for Constructivist strategies compared to Behaviorist strategies observed in the present study should thus not be construed as a commentary on the relative value of Constructivist instructional practices, but rather a reflection of the lack of alignment with the outcome measures used.

Over against the results for Constructivist instructional strategies is the comparatively strong showing for traditional, Behaviorist instructional methods. From the tenor of many recent comparisons of instructional practices, “traditional” instruction would appear to be a far inferior instructional method practiced by educators dwelling in some proverbial instructional dark age. The results of this meta-analysis—at least in respect to Web-based delivery of instruction—belie that impression. Contrary to the picture of old fashioned ineffectiveness often implied or stated outright concerning traditional higher instructional practices, such practices appear, in fact, to be quite effective at doing what they are good doing: delivering a large amount of instruction to a large number of students in the least amount of time with reasonable levels of achievement. That this is true should take no more than simple reflection on the sheer numbers of college and university graduates produced in any given year, not to mention the sheer cumulative numbers of such graduates extant worldwide—all meeting at least minimum standards of quality.

What seems a more likely explanation is that Behaviorism and Constructivism are two different instructional methods with different instructional purposes and each is best suited for—and measured by—different instructional situations. Shield (2000) points out that Behaviorist instructional practices are especially adept at laying the informational background necessary for more in-depth learning—which is why they remain relevant today. In contrast, Constructivist instruction seems particularly well-suited to exploring narrow topics in greater depth leading to increased understanding. Thus, Behaviorism and Constructivism are not antithetical, but complementary. Perhaps this complementarity may be best observed in the difference between undergraduate and graduate education. Graduate education relies upon students entering with basic background knowledge already in place and focuses on greater in-depth understanding of selected portions of that background knowledge. Graduate level instruction is narrower and more in-depth than undergraduate instruction and post-graduate instruction even more so: learning gets increasingly deep and increasingly narrow as one progresses and the instructional activities and strategies change with the progression. In the current study, 9 of the 17 studies that form the Constructivist group involved undergraduate students. Investigating the best fit for Constructivist versus Behaviorist instructional strategies appears to be profitable avenue for future research. As stated earlier, caution should be used when considering the results of this exploratory study as a basis for implementing instructional practices.

Media Delivery. Due to the coding rules adopted in this meta-analysis, the category "media delivery" is primarily composed of instructional situations where no instructional activity or strategy is identified as a contrast. The prevailing intention on the part of the researcher was to create instructional equivalence between the f2f and the on-line condition. Thus, the on-line instruction was deliberately designed to emulate f2f instruction as closely as possible. In those cases where studies were identified as media delivery only and the instructional activities were identified, those activities were invariably behaviorist in nature. Thus, media delivery can be considered a type of Behaviorist instructional strategy where no specific instructional activity was featured as a contrast. In most cases in the studies coded as Behaviorist, some specific instructional activity—based on a particular instructional theory—was being added to the treatment condition while all other instruction between the treatment and contrast was kept equivalent. Thus, there is good reason to believe that the behaviorist and media delivery categories differ primarily by the addition of a specific instructional activity to otherwise behaviorist instructional strategies. The natural outcome of this is that the Behaviorist category should be more effective than the media delivery category simply because students in the Behaviorist category were exposed to additional instructional interventions. That fails, however, to explain the rather large difference in effect size between Behaviorist and media delivery. If the difference between the two is primarily due to the addition of one instructional activity, is it possible that one instructional activity—regardless of what it is—can make such a dramatic difference?

Post hoc tests for moderator shed some light on this question. According to tests for the effect of instructor role, it was found that the impact of formative feedback (instructor role) was almost twice as large on student achievement in Behaviorist designs as in media delivery studies. It is hypothesized that the difference in outcomes between Behaviorist and media delivery studies is a function of two factors: the addition of one or more instructional activities and the presence of increased levels of instructor-based formative feedback in studies coded as Behaviorist versus those coded as media delivery only.

Collaborative Design

Consistent with the finding that Behaviorist instructional strategies appear to be very effective when used in Web-based distance education, is the related finding that the most effective collaborative design is required collaboration that is moderated by the instructor. That finding is accompanied by the even more surprising result that no collaboration whatsoever appears to be more effective than collaboration in which students take the leadership role. The implication that could easily be drawn from these results is that collaboration is less effective when students are left to their own devices than when it is directed by an instructor—which conflicts with the results observed earlier that suggest that students learn more effectively on their own than with an instructor.

One interpretation of this might be that collaboration itself interferes with certain types of learning and the increased time required for discussion and negotiations required to reach consensus are counterproductive for certain types of learning. Like Constructivist Strategies, Collaborative Design may be similarly sensitive to both the purposes and the methods used to measure the outcomes of learning. Thus, in the wrong situation, using collaboration simply for the sake of using collaboration may, in fact, detract from learning rather than enhance it. To the degree that Web-based instruction forms an instantiation of a particular type of instructional situation, it appears that student-facilitated or voluntary collaboration is far less effective as an instructional strategy than either no collaboration or teacher-moderated collaboration. This observation is limited purely to the effect on reported achievement outcomes and it is acknowledged that collaboration may be an effective tool for increasing the retention rate—which would likely have an indirect effect on the achievement outcomes.

The most likely theoretical explanation for the findings regarding collaboration reported here is that teacher-led collaboration is more content-centered than is student-led collaboration. That is, while student-led collaboration may improve student attitudes toward a course, at the same time it may be more of a distracter than an aid to interaction with content. Alternatively, some types of collaboration, like some instructional strategies, may not be as amenable to objective measures of achievement as others.

Summary, Conclusions, Implications

Summary

The aggregate effect size reported in this meta-analysis suggests that web-based DE is a highly effective method for delivery of instruction to adult students. That effectiveness also appears to have increased over time. In addition, instructional designs that favor student-content interactions appear to be somewhat more effective than designs that favor interpersonal interactions. Collaborative designs in which the instructor acts as moderator appear to be more effective than other collaborative designs and instructor provision of formative feedback was the most significant modifier of effect size observed in this meta-analysis. Multimedia delivery, simulations and problem-based learning all emerged as having more effect on student outcomes than other instructional activities examined.

The above findings are based upon a representative and robust sample and are generalizable to the sample population of adult learners using Web-based instruction. That generalization is limited by the use of only one coder and the highly heterogeneous nature of the included sample studies.

Conclusions

In much that same way that qualitative research and quantitative research complement each other, so, apparently, do teacher-centered and student-centered instructional strategies. The traditional, somewhat adversarial relationship between teacher-centered (i.e., traditional behaviorist) and student-centered (i.e., constructivist) seems ill-advised and counter-productive. Instead, research might best focus on identifying those situations in which a particular set of instructional strategies is most appropriate for the purpose of that instructional situation. For instance, introductory courses wherein the primary purpose is to provide background knowledge (i.e., the body of knowledge and or skills to be learned is already known and identified) for more advanced learning may benefit more from traditional, behaviorist instructional strategies (lecture, reading, some discussion, teacher-directed activities and simple drill and practice and memorization) than from student-centered instruction, particularly if course assessments emphasize discrete knowledge measured by objective tests. Conversely, in advanced studies instruction where a certain level of background knowledge, skill or experience is assumed on the part of all learners, student-centered constructivist strategies that lead to in-depth exploration of a few topics may be more appropriate.

Implications for Research

One of the most pressing needs for future primary research highlighted by this study is the investigation of the alignment of outcomes measures with instructional strategies, particularly for use with Constructivist instruction. Likewise, research on the alignment between various types of collaboration and outcome measures is also needed. Future meta-analyses of Web-based DE will be dependent upon additional primary research directly contrasting specific instructional activities with others.

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

*Adcock , A. B., Duggan, M. H., Nelson, E. K., Nickel, C. (2006). Teaching effective helping skills at a distance: The development of Project CATHIE. *The Quarterly Review of Distance Education, 7*(4), 349–360.

Admiraal, W., Wubbels, T., and Pilot, A. (1998). College Teaching in Legal Education: Teaching Method, Students' Time-on-Task, and Achievement. *Research in Higher Education*, 40(6), 687–704.

Ahern, T. C. & Durrington, V. (1995). Effects of anonymity and group saliency on participation and interaction in a computer-mediated small group discussion. *Journal of Research on Computing in Education, 28*, 133-147.

Akkuş, H., Kadayifçi, H., Atasoy, B. & Geban, Ö. (2003). Effectiveness of instruction based on the constructivist approach on understanding chemical equilibrium concepts. *Research in Science and Technology Education*, *21*(2), 209-227.

*Alavi, M., Marakas, G. M., & Yoo, Y. (2002, Dec). A Comparative Study of distributed learning environment on learning outcomes. *Information Systems Research, 13*(4), 404-415.

Althaus, S. L. (1997). Computer-mediated communication in the university classroom: An experiment with on-line discussions. *Communication Education, 46*(3), 158-174.

American Council on Education. (2001). *Distributed education: Challenges, choices and a new environment*. Washington, D.C.: American Council on education/EDUCAUSE.

American Psychological Association (APA). (2010). *Publication manual of the American Psychological Association*. Washington, D.C.: Author.

Anderson, J. A. (1988). Cognitive styles and multicultural populations. *Journal of Teacher Education*, *39*(1), 2-9.

*Anderton, E. K. (2005). *An evaluation of strategies to promote self-regulated learning in pre-service teachers in an on-line class*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3193866)

Archambault, L., & Crippen, K. (2009). Examining TPACK among K-12 online distance educators in the United States: contemporary issues in technology and teacher education. *CITE Journal, 9*(1). Available from: http://www.citejournal.org

Aronson, J., Zimmerman, J. & Carlos, L. (1999). *Improving student achievement by extending school: Is it just a matter of time?* San Francisco: WestEd. [ERIC: ED 435 127] http://web.wested.org/online_pubs/timeandlearning/1_intro.html

* Banks, L. V. (2004). *Brick, click, or crick and click: A comparative study on the effectiveness of content delivery modalities for working adults*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3122174)

Barr, R. B. & Tagg, J. (1995). From teaching to learning—A new paradigm for undergraduate education. *Change, 2*(5), 13-25.

Bartlett, II, J. E., Reynolds, K. A. & Alexander, M. W. (2000). A tool for on-line learning. *Journal of Online Learning, 11*(3-4), 22-24.

*Baturay, M. H. & Bay, O. F. (2010). The effects of problem-based learning on the Classroom community perceptions and achievement of web-based education students. *Computers & Education*, *55,* 43–52. doi: 10.1016/j.compedu.2009.12.001

Bax, L., Yu, L-M., Ikeda, N. & Moons. K. G. M. (2007). A systematic comparison of software dedicated to meta-analysis of causal studies. *BMC Medical Research Methodology*, *7,* 1-21. Available from: http://www.pubmedcentral.nih.gov

Becker, W. E. & Watts, M. (2001). Teaching methods in U.S. undergraduate economics courses. *Journal of Economic Education, 32*(3), 269-279.

Bee, R. H. & Usip, E. E. (1998). Differing attitudes of economics students about Web-based instruction. *College Student Journal, 32*(2), 258-269.

Beer, M., Slack, F, & Armitt, G. (2005). Collaboration and teamwork: Immersion and presence in an on-line learning environment. *Information Systems Frontiers, 7*(1), 27-37.

Begg C.B. & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1088-101.

Bell, T., Schrum, L., & Thompson, A. D. (eds.) (2008). *Framing research on technology and student learning in the content areas: Implications for educators*. Charlotte, NC: Information Age Publishing.

* Benjamin, S. E., Tate, D. F., Bangdiwala, S. I., Neelon, A. S., Ammerman, J. M., Dodds, B. H., & Ward, D.S. (2008). Preparing child care health consultants to address childhood overweight: A randomized controlled trial comparing Web to in-person training. *Maternal and Child Health Journal,* 12(5), 662-669. doi: 10.1007/s10995-007-0277-1

Benson, A., Johnson, S., Taylor, G. Treat, T., Shinkareva, O. & Duncan, J. (2005). Achievement in online and campus-based career and technical education (CTE) courses. *Community College Journal of Research and Practice, 29*(5), 369-394.

Berge, Z. L. (1999). Interaction in postsecondary Web-based learning. *Educational Technology*, *1*, 5-10.

Bernard, R. M. & Abrami, P. C. (2009). *Meta-analysis of distance education and online learning research: Past, present and future*. Invited presentation at the 15th annual Sloan-C International Conference on Online Learning, 28-30 October 2009, Orlando, FL.

Bernard, R. M., Abrami, P. C., Lou, Y., & Borohovski, E. (2004a). A methodological morass? How we can improve the quality of quantitative research in distance education. *Distance Education*, *25*(2), 176-198.

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., Wallet, P. W., Fiset, M. & Huang, B. (2004b, Fall). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research, 74*(3), 379-439.

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A. & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research, 79*(3), 1243-1289.

*Bernard, R. M. & Lundgren-Cayrol, K. (2001, Sept). Computer conferencing: An Environment for collaborative project-based learning in distance education. *Educational Research & Evaluation*, *7*(2/3), 241-261.

Bielman, V., Putney, L. & Strudler, N. (2000). *Constructing community in a postsecondary virtual classroom*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

BioStat. (2009). Comprehensive Meta-Analysis (CMA) (Version 2.2). [Computer software]. Englewood, NJ: Biostat.

* Bixler, B. A. (2008/2007). *The effects of scaffolding student's problem-solving process via question prompts on problem solving and intrinsic motivation in an online learning environment*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3284910)

Black, S. (2003, Dec.). Engaging the Disengaged. *The American School Board Journal,* 190(12), 58–60.

Blignaut, A. S. & Trollip, S. R. (2003). Measuring faculty participation in asynchronous discussion forums. *Journal of Education for Business*, *78*(6), 347-353.

Bocchi, J., Eastman, J. K., & Swift, C. O. (2004). Retaining the online learner: Profile of students in an online MBA program and implications for teaching them. *Journal of Education for Business*, *79*(4), 245-253.

Bodzin, A.M., & Park, J.C. (2000). Dialogue patterns of preservice science teachers using asynchronous computer-mediated communications on the World Wide Web. *Journal of Computers in Mathematics and Science Teaching, 19*(2), 161-194.

Bohlin, R. M. (1997). Computer technology can be the irresistible force needed to move education. *Renaissance, 4*(1), 14-16.

* Boulter, M. L. (2010). *The influence of Socratic Questioning in online discussions on the critical thinking skills of undergraduate students: An exploratory study based on a business course at a proprietary university.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3397620)

Boyle, S. L. T., Kolosh, K., L'Allier, J. & Lambrecht, J. (2003). Thompson's NETg's blended-learning model: The next generation of corporate and school-based learning. *The Delta Pi Epsilon, 45*(3), 145-161.

Braman, O. R. (1998). *The cultural dimension of individualism and collectivism as a factor in adult self-directed learning readiness*. (Doctoral dissertation). University of Southern Mississippi.

Brophy, J. E. (1988). Educating teachers about managing classrooms and students. *Teaching and Teacher Education, 4*, 1, 3.

Brown, B. L. (2003). *Teaching style vs. learning style. Myths and realities*. ERIC Clearinghouse on Adult, Career and Vocational Education. No. 26.

Brown, J. S., & Duguid, P. (2000). *The social life of information*. Boston: Harvard Business School Press.

Byrd, J. K. (2001). *Student achievement: Is equality really necessary?* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, Washington, April 10-14, 2001. [ERIC Document Retrieval No.: ED 452 273.]

* Caldwell, E. R. (2006). A comparative study of three instructional modalities in a computer programming course: Traditional instruction, web-based instruction, and online instruction. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3227694)

Campbell, M., Gibson, W., Hall, A., Richards, D. & Callery, P. (2008). Online vs. face-to-face discussion in a Web-based research methods course for postgraduate nursing students: A quasi-experimental study. *International Journal of Nursing Studies, 45*(5), 750-759.

Cardwell, D. (1995). *Wheels, clocks, and rockets: A history of technology*. New York: Norton.

Carroll, J. (1963). A model of school learning. *Teachers College Record, 64*, 723-733.

Case, J., Gunstone, R. & Lewis, A. (2000, April/May). *Approaches to learning in a second year chemical engineering course*. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, New Orleans, LA.

* Castaneda, R. (2008). *The impact of computer-bsed simulations within an instructional sequence on learner performance*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3304816)

Cavanaugh, C. S. (2001). The effectiveness of interactive distance education technologies in K-12 learning: A meta-analysis. *International Journal of Educational Telecommunications, 7*(1), 73-88.

* 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*(3), 301-321. [Cited as Cavus, Uzonboylu & Ibrahim 2007 by Means, et al. (2009)]

* Chang, S. L. & Chang, Y. (2008). Using online concept mapping with peer learning to enhance concept application. *The Quarterly Review of Distance Education, 9*(1), 17-27.

Charp, S. (2002, April). Editorial: Educator's acceptance of computer technology? *T.H.E. Journal (Technological Horizons in Education)*, *29*(9), 10-12.

* Chen, B., Hirumi, A. & Zhang, N. J. (2007). Investigating the use of advance organizers as an instructional strategy for web-based distance education. *The Quarterly Review of Distance Education, 8*(3), 223–231.

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

Chickering, A. W. & Gamson, Z. F. (1987, March). Seven principles for practice in undergraduate education. *AAHE Bulletin*, 3–7.

Choi, H. J., & Johnson, S. D. (2005). The effect of context-based video instruction on learning and motivation in online courses. *The American Journal of Distance Education, 19*(4), 215-227.

Christel, M. G. (1994). The role of visual fidelity in computer-based instruction. *Human-Computer Interaction, 9*, 183-223.

Christopher, M. M., Thomas, J. A., & Tallent-Runnels, M. K. (2004). Raising the bar: encouraging high level thinking in online discussion forums. *Roeper Review, 26*(3), 166-171.

* Clapano, K. (2010). *The effects of content presentation and learning styles on recall in an online course environment*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3437244).

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*(2), 21-29.

Clark, R E. (2000). Evaluating distance education: Strategies and cautions. *Quarterly Review of Distance Education, 1*, 3-16.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10,* 101–129.

Cochrane Collaboration. (2000). The Cochrane reviewers handbook glossary, version 4.1.5 retrieved December 6, 2002 from http://www.cochrane.de/software/Documentation/Handbook/glossary.doc

Coeyman, M. (2002, July). Popular reform draws mixed reviews. *Christian Science Monitor, 13*. Available on-line at: http://www.csmonitor.com

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. San Diego, CA: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

*Collins, M. (2000). Comparing Web, correspondence, and lecture versions of a second-year non-major biology course. *British Journal of Educational Technology, 31*(1), 21-27.

Colon, T., & Simpson, M. (2003). Silicon Valley versus Silicon Glen: the impact of computers upon teaching and learning: A comparative study. *British Journal of Educational Technology*, *34*(2), 137-150.

* Connolly, T. M., MacArthur, E., Stansfield, M. & McLellan, E. (2007). Quasi-experimental study of three online learning courses in computing. *Computers & Education 49,* 345–359. doi:10.1016/j.compedu.2005.09.001

* Cook, D. A., Gelua, M. H., Dupras, D. M., & Schwartz, A. (2007). Instructional methods and cognitive and learning styles in web-based learning: report of two randomized trials. *Medical Education, 41*, 897-905. doi:10.1111/j.1365-2923.2007.02822.x

Cooper, H.M. (1979). Statistically combining independent studies: A meta-analysis of sex differences in conformity research. *Journal of Personality and Social Psychology*, 37, 131–146.

Cooper, H., & Hedges, L. V., (Eds.). (1994). *The handbook of research synthesi*s. New York: The Russell Sage Foundation.

Cooper, L. W. (2001). A comparison of online and traditional computer applications classes. *T.H.E. Journal, 28*(8), 52-58.

Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Cambridge, MA: Harvard University Press.

Davis, S. & Meyer, C. (1998). *Blur: The speed of change in the connected economy*. Reading, MA: Addison-Wesley.

Desai, M., Hart, J., & Richards, T. (2008). E-learning: paradigm shift in education. *Education*, *129* (2), 327-334.

Dick, W.& Carey, L. (1985). *The systematic design of instruction*. Glenview, IL: Pearson, Scott, Foresman.

Doughtery, O. J. & Done, D. J. (2009). Is semantic memory impaired in schizophrenia? A systematic review and meta-analysis of 91 studies. *Cognitive Neuropsychiatry, 14*(6), 473-509. doi: 10.1080/13546800903073291

* Draper, D. C. (2010). *The instructional effects of knowledge-based community of practice learning environments on student achievement and knowledge convergence*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3436066).

Duval, S. & Tweedie, R. (2000). Trim and fill: a simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *50***,** 455–463.

Egger, M, Smith, G. D., Schneider, M. & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629-634.

Entwistle, N. (2001). Styles of learning and approaches to studying in higher education. *Kybernetes, 30*(5/6), 593-604.

Entwistle N. & Tait H. (1990). Approaches to learning, evaluations of teaching, and preferences for contrasting academic environments. *Higher Education 19,* 169-194.

Eppler, M. A. & Ironsith, M. (2004). PSI and distance learning in a developmental psychology course. *Teaching of Psychology, 31*(2), 131-134.

Farmer, L. S. J. (2005). Using technology to facilitate assessment of library education. *Teacher Librarian, 32*(2), 12-15.

Figueroa, M. L. (1992). Understanding students' approaches to learning in university traditional and distance education courses. *Journal of Distance Education, 7*(3), 15-26.

Flavell, J. H. (1979, Oct.). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), p906-11.

Flint, W. J., (2004). *Faculty development toward a learning college: Peer mentoring and critical reflection on learner-centered teaching techniques*. (Doctoral dissertation, Capella University, 2004). Retrieved from ProQuest database January 28, 2011. (UMI No. 3138938.)

* Fox, K. D. (2010). *Investigating the impact of multimedia design principles on learning in an online context*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3397444).

Fram, A. & Tompson, T. (2007). Teens: Don't confront, simply text. Associated Press via *The Hanford Sentinel*, Thursday, November 15, 2007, p. 1

Frey, A., Faul, A., & Yankelov, P. (2003). Student perceptions of Web-assisted teaching strategies. *Journal of Social Work Education, 39*(3), 443-457.

* Frey, T. (2008). Determining the impact of online practicum facilitation for inservice teachers. *Journal of Technology and Teacher Education, 16*(2), 181- 210.

*Frith, K. H. & Kee, C. C. (2003). The effect of communication on nursing student outcomes in a web-based course. *Journal of Nursing Education, 42*(8), 350-358.

Gagne, R. M. & Briggs, L. J. (1979). *Principles of instructional design*. New York: Holt, Rinehart and Winston.

Gayton, J. & McEwen, B. C. (2007). Effective online instructional and assessment strategies. *American Journal of Distance Education, 21*(3), 117-132.

Gilbert, P. K., & Dabbagh, N. (2005). How to structure online discussions for meaningful discourse: A case study. *British Journal of Educational Technology, 36*(1), 5−18.

Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher, 10*, 3-8.

Glass, G. V. (1978a). Reply to Mansfield and Busse. *Educational Researcher, 5*, 3-8.

Glass, G. V. (1978b). Integrating findings: The meta-analysis of research. *Review of Research in Education, 5*, 351-379.

Glass, G. V., McGaw, B., & Smith, M.L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Greene, B. A. & Land, S. M. (2000). A qualitative analysis of scaffolding use in a resource-Based learning environment involving the World Wide Web. *Journal of Educational Computing Research, 23*(2), 151.

Greene, B., Lewis, L., Snow, K., Farris, E., & Levin, D. (1999). *Distance education at postsecondary education institutions: 1997-98.* NCES 2000-013Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Gross, K. & Godwin, P. (2005). *Education's many stakeholders: Educational administrators are increasingly recognizing what businesses have long understood: customer satisfaction matters.* Online http://www.universitybusiness.com/viewarticlepf.aspx?articleid=174

* Gulikers, J. T. M., Bastiaens, T. J. & Martens, R. L. (2005). The surplus value of an authentic learning environment. *Computers in Human Behavior, 21*, 509–521. doi:10.1016/j.chb.2004.10.028

* Gupta, S. (2006). *Longitudinal investigation of collaborative e-learning in an end-user raining context*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3208751).

* Hairston, N. R. (2007). Employee's Attitude toward e-learning: implications for policy in industry environments. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3257874)

* Hansen, B. A. (2000). *Increasing person-environment fit as a function to increase adult learner success rates in distance education*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9993734)

* Hansen, D. E. (2008). Knowledge transfer in online learning environments. *Journal of Marketing Education, 30*(2), 93-105. doi: 10.1177/0273475308317702

Hansen, N. E.& Gladfelter, J. (1996). Teaching graduate psychology seminars using electronic mail: Create distance education. *Teaching of Psychology, 23*(4), 252-256.

Hartman, H. J. (2001). *Metacognition in Learning and Instruction: Theory, Research and Practice*. Dordrecht: Kluwer Academic Publishers.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128. doi: 10.3102/10769986006002107.

Hedges, L. V. (1994a). Statistical considerations. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 29-38). New York: Russell Sage Foundation.

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

Hedges, L. V. & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods, 9*(4), 426-445.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21,* 1539–1558.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327,* 557–560.

Holmes Jr., O. W. (1897). The path of the law. *Harvard Law Review,* 10, 457. Available at: http://www.constitution.org/lrev/owh/path_law.htm

Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006, June). Assessing heterogeneity in meta-analysis: Q statistic or I[squared] index? *Psychological Methods, 11*(2), 193-206. ERIC [EJ747791]

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage.

* Hylton, M. E. (2006). Online versus classroom-based instruction: A comparative study of learning outcomes within a diversity course. *Journal of Baccalaureate Social Work*, *11*(2), 102-114.

Im, Y. & Lee, O. (2003/2004). Pedagogical implications of online discussion for preservice teacher training. *Journal of Research on Technology in Education, 36*, 155-170.

* Isenberg, N. A. (2010). *A comparative study of developmental outcomes in webbased and classroom-based German language education at the post-secondary level: vocabulary, grammar, language processing, and oral proficiency development*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3420155).

Jackman, D. H. & Swan, M. K. (1996). *Instructional models effective in distance education*. Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 8-12, 1996. [North Dakota State University]

Jackson, S. & Wolski, S. (2001). Identification of and adaptation to students' preinstructional beliefs in introductory communication research methods: contributions of interactive web technology. *Communication Education, 50*(3), 189-205.

Jahng, N., Krug, D. & Zhang,Z. (2007). Student achievement in online distance education compared to face-to-face education. *European Journal of Open, Distance and E-learning*. Retrieved from http://www.eurodl.org/materials/contrib/2007/Jahng_Krug_Zhang.htm 1/27/2010.

* Jang, K. S., Hwang, S. Y., Park, S. J., Kim, Y. M. & Kim, M. J. (2005, Jan.). Effects of a web-based teaching method on undergraduate nursing students' learning of electrocardiography. *Journal of Nursing Education, 44*(1), 35-39.

Jelf, A. & Colburn, C. (2002). Do students' approaches to learning affect their perceptions of using computing and information technology? *Journal of Educational Media, 27*(1-2), 41-53.

Johnson, D. W., Johnson, R. T. & Stanne, M. E. (2000). *Cooperative learning methods: a meta-analysis*. Minneapolis, MN: University of Minnesota Press.

Jonassen, D. H. (1996). *Computers in the classroom: Mindtools for critical thinking*. Englewood Cliffs, NJ: Merrill.

Jonassen, D. H. (2000a). Toward a design of a problem-solving theory. *Educational Technology Research & Development, 48*(4), 63-85.

Jonassen, D. H. (2000b). *Computers as Mindtools for schools: Engaging in critical thinking,* (2nd ed.). Upper Saddle River, NJ: Merrill.

Jonassen, D. H. (2000c). Transforming learning with technology: Beyond Modernism and Postmodernism or whoever controls the technology creates the reality. *Educational Technology, 40*(2), 21-25.

Jonassen, D. & Land, S. M. (eds.). (2000). Preface. In Jonassen, D. & Land, S. M. (eds.), *Theoretical foundations of learning environments*. Mahweh, NJ: Erlbaum.

Jonassen, D., Davidson, M., Collins, A., Campbell, J., & Bannan-Haag, B. (1995). Constructivism and computer-mediated communication in distance education. *The American Journal of Distance Education, 9*(2), 7-26.

Joyce, B., Weil, M. & Showers, B. (1992). *Models of teaching*. Boston: Allyn and Bacon.

* Jung, I., Choi, S., Lim, C. & Leem, J. (2002). Effects of different types of interaction on learning achievement, satisfaction and participation in web-based instruction. *Innovations in Education and Teaching International, 39*(2), 153-162. Doi: 10.1080/1355800021012139 9

Kaestle, C. (1993). The awful reputation of education research. *Educational Researcher, 22*(1), 23-31.

* Kanuka, H. & Jugdev, K. (2006, June). Distance education MBA students: an investigation into the use of an orientation course to address academic and social integration issues. *Open Learning, 21*(2), 153–166. doi: 0.1080/02680510600715578

* Karatas, S. & Simsek, N. (2009). Comparisons of internet-based and face-to-face learning systems based on “equivalency of experiences” according to students’ academic achievements and satisfactions. *The Quarterly Review of Distance Education, 10*(1), 2009, 65–74.

* Karr, C. L., Weck, B., Sunal, D. W. & Cook, T. M. (2003). Analysis of the effectiveness of online learning in a graduate engineering math course. *The Journal of Interactive Online Learning, 1*(3).

Kaye, A. (1992). Learning together apart. In A.R. Kaye (Ed.), *Collaborative Learning Through Computer Conferencing: The Najaden Papers* (pp. 1-24). New York: Springer.

Keefe, T. J. (2003). Using technology to enhance a course: the importance of interaction. *EDUCAUSE Quarterly, 1*, 24-34.

Kember, D. & Gow, L. (1994). Orientations to teaching and their effect on the quality of student learning. *Journal of Higher Education, 65*, 59–74.

* Kemper, K. J., Gardiner, P., Gobble, J., Mitra, A., & Woods, C. (2006). Randomized controlled trial comparing four strategies for delivering e-curriculum to healthcare professionals. *BMC Medical Education, 6*(2), 1-11. doi: 10.1186/1472-6920-6-2

Knowles, M. S. (1980). *The modem practice of adult education: From pedagogy to andragogy*. Chicago: AP/Follett.

Knupfer, N. N., Gram, T. E. & Larsen, E. Z. (1997). *Participant analysis of a multi-class, multi-state, on-line, discussion list.* (ERIC Document Reproduction Service No. ED 409-845).

* Krall, R. M., et al. (2009). Hands-on at a distance: evaluation of a temperature and heat distance learning course. *Journal of Science Education and Technology, 18*, 173-186. doi: 10.1007/s10956-008-9142-2

Kulik, J., Kulik, C., & Cohen, P. (1980). Effectiveness of computer-based college teaching: A meta-analysis of findings. *Review of Educational Research, 2*(2), 525-544.

Lant, K. M. (2002). *Flesh and bone: information literacy, teaching and the connected classroom*. Paper presented at the Annual Mid-south Instructional Technology Conference. (ERIC Document Reproduction Service No. ED464624).

* LaRose, R., Gregg, J., & Eastin, M. (1998). Audiographic telecourses for the Web: An experiment. *Journal of Computer Mediated Communication, 4*(2). Available on-line at: http://jcmc.indiana.edu/vol4/issue2/larose.html

Lasswell, H.D. (1971). *Propaganda technique in the World War*. Boston: The MIT Press.

* Lee, Z-H. (2010). *An experimental study on situated and dynamic learning assessment (SDLA) environment*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3417762).

Lepp, G. A. (2010). *Knowledge and use of student-centered instructional techniques, alternative assessment methods, and new educational technology in adult business education*. (Doctoral Dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3409187).

Levin, S. R., Levin, J.A. & Chandler, M. (2001). *Social and organizational factors in creating and maintaining effective online learning environments*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Levin, J., Levin, S. R. & Waddoups, G. (1999). Multiplicity in learning and teaching: A framework for developing innovative online education. *Journal of Research on Computing in Education, 32*, 256-268.

Liang, X., & Creasy, K. (2004). Classroom assessment in Web-based instructional environment: Instructors'experience. *Practical Assessment, Research & Evaluation, 9*(7). Available on-line at http://PAREonline.net/getvn.asp?v=9&n=7

Long, D. R. (1983, Nov). *Time-on-task in beginning students of Spanish at the university level: A case study*. Paper presented at the Annual Meeting of the American Council on the Teaching of Foreign Languages, San Francisco. ERIC Document Reproduction Service No. ED 239 514.

Long, H. B. (1994). Challenging some myths about Self-Directed Learning. In H. B. Long, et al., *New Ideas about Self-Directed Learning*, pp. 1-14. Norman: University of Oklahoma.

Lord, T. R. (1997). A comparison between traditional and constructivist teaching in college biology. *Innovative Higher Education,* 21, 197–216.

Long, M. & Jennings, H. (2005). *"Does it work?" The Impact of technology and professional development on student achievement*. Calverton, MD: Macro International.

Lou, Y., Bernard, R. M. & Abrami, P. C. (2006). Media and pedagogy in undergraduate distance education: a theory-based meta-analysis of empirical literature. *Educational Technology Research and Development, 54*(2), 141-176.

Lyons, L. C. (1998). *Meta-analysis: Methods of accumulating results across research domains*. Available at: http://www.lyonsmorris.com/MetaAnalysis.htm

McGraw, K. O. & Wong, S. P. (1992). A common effect size statistic. *Psychological Bulletin, 111*(2), 361-365.

McIssac, M. S., Blocher, J. M., Mahes, V. & Vrasidas, C. (1999). Student and teacher perception of interaction in on-line computer-mediated communication. *Educational Media International, 36,* 121-131.

McLuhan, M. (1994). *Understanding media: The extensions of man*. Cambridge, MA: MIT Press.

McShannon, J., Hynes, P., Nirmalakhandan, N., Venkatatamana, G., Ricketts, C., Ulery, A., & Steiner, R., (2006). Gaining retention and achievement for students program: A faculty development program. *Journal of Professional Issues in Engineering Education and Practice, 132*(3), 204-208.

Maddrell, J. (2008). Interaction in computer mediated distance education. Unpublished paper available via Scribd (Facebook document sharing), No. 3728783.

Maehl, W. H. (2000). *Lifelong learning at its best.* San Francisco: Jossey-Bass.

Maki, R. Maki, W. Patterson, M. & Whittaker, D. (2000). Evaluation of a Web-based introductory psychology course: I: Learning and satisfaction in online versus lecture courses. *Behavior Research Methods, Instruments and Computers, 32(*2), 230-239.

Marzano, R. J. (1998). *A theory-based meta-analysis of research on instruction*. A report sponsored by the Office of Educational Research and Improvement, United States Department of Education. Aurora, CO: Mid-Continent Regional Educational Laboratory.

Mayer, D. (2010). *Essential evidence-based medicine, second edition*. Cambridge, NY: Cambridge University Press.

Mayer, R. E. & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal ofEducational Psychology, 93*, 390-397.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, *93*, 187-198.

Mayers, T., & Swafford, K. (1998). Reading the networks of power. In T. Taylor, & I Ward (Eds.), *Literacy theory in the age of the Internet* (pp. 146-157). New York: Columbia University Press.

Means, B., & Haertel, G. D. (2004). *Using technology evaluation to enhance student learning*. New York: Teacher's College Press.

Means, B., Toyama, Y., Murphy, R., Bakia, M. & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online]LTRNGG learning studies*. A report prepared for the U.S. Department of Education, Office of Planning, Evaluation and Policy Development. Contract number ED-04-CO-0040 Task 0006. Washington, D.C.

* Mebane, M. Porcelli, R., Iannone, A., Attanasio, C. & Francescato, D. (2008). Evaluation of the efficacy of affective education online training in promoting academic and professional learning and social capital. *International Journal of Human-Computer Interaction*, *24*(1), 68-86. doi: 10.1080/10447310701771498

Miller, K. J. & Vanni, M. (2005). Inter-rater agreement measures, and the refinement of metrics in the PLATO MT evaluation paradigm. *MT Summit X, Phuket, Thailand, September 13-15, 2005, Conference Proceedings: The tenth Machine Translation Summit*; pp.125-132. Available from: http://www.mt-archive.info/MTS-2005-Miller.pdf

Moore, M. (1989). Three types of interaction [Electronic version]. *The American Journal of Distance Education, 19*(3), 133. Retrieved from http://www.ajde.com/Contents/vol3_3#editorial

Moore, D.M. (1994). The parable of the expensive ballpoint pen (revisited): Implication for hypermedia. *Computers in the Schools*, *10*(1/2), 3-7.

Nickel, C. E. (2010). The effects of cooperative and collaborative strategies on student achievement and satisfaction in blended and online learning environments. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3416872).

Nussbaum, E. M., Hartley, K., Sinatra, G. M., Reynolds, R. E. & Bendixen, L. D. (2004). Personality interactions and scaffolding in on-line discussions. *Journal of Educational Computing Research, 30*(1 & 2), 113-137.

O'Donnell, D. (1999). Habermas, critical theory and selves-directed learning. *Journal of European Industrial Training, 23*(4-5), 251-261.

Ong, W. J. (1958). *Ramus, method and the decay of dialogue*. Cambridge, MA: Harvard University Press.

Oppenheimer, T. (1997, July). The computer delusion, *The Atlantic Monthly*, 45-48, 50-56, 61-62.

O'Reilly, T. (2000, Winter). Everything in connected: Ubiquitous computing will turn the computing world topsy turvy. *The Future of Software Magazine, 1*(1), 72-73.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8*, 157-159.

Orwin, R. G. (1994). Evaluating coding decisions. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 139-162). New York: Russell Sage Foundation.

* Own, Z. (2006). The application of an adaptive web-based learning environment on oxidation-reduction reactions. *International Journal of Science and Mathematics Education, 4*, 73–96.

* Pacifici, C., et al (2006). Web-based training for foster, adoptive, and kinship parents. *Children and Youth Services Review*, *28*, 1329–1343. doi:10.1016/j.childyouth.2006.02.003

Papert, S. (1980). *Mindstorms. Children, computers and powerful ideas*. New York: Basic Books.

Parsad, B., & Lewis, L. (2008). *Distance education at degree-granting postsecondary institutions: 2006–07* (NCES 2009–044). Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

* Parsons, R. (2006). *Improving preceptor self-efficacy using an on-line educational program*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3225762)

Patterson, D. A. (1996). Microprocessors in 2020. In *Key technologies for the 21st century: A Scientific American special issue* (pp.1-8). New York: W. H. Freeman and Company.

Peddiwell, J. A. (1939/2004). *The saber-tooth curriculum*. New York: McGraw Hill.

* Peterson, C. L. & Bond, N. (2004, Summer). Online compared to face-to-face teacher preparation for learning standards-based planning skills. *Journal of Research on Technology in Education, 36*(4), 345-360.

Phipps, R. & Merisotis, J. (1999). *What's the difference? A review of contemporary research on the effectiveness of distance learning in higher education*. (Report No. HE 032019). Washington, D. C.: Institute for Higher Education Policy. (ERIC Document Reproduction Service No. ED 429 524).

Pintrich, P. R. (1988). A Process-oriented view of student motivation and cognition. In Stark, J., & Mets, L. (Eds.). *Improving Teaching and Learning Through Research*, 65–79. San Francisco, CA: Jossey-Bass.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385-407.

Poirier, C. R. & Feldman, R. S. (2004). Teaching in cyberspace: online versus traditional instruction using a waiting-list experimental design. *Teaching of Psychology, 31*(1), 59-62.

Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301-322). New York: Russell Sage Foundation.

Resnick, M. (1994). *Turtles, termites, and traffic jams: Explorations in massively parallel microworlds*. Cambridge, MA: MIT Press.

Roberts, R. M. (2002). *The role of computers in school restructuring: a meta-analysis*. Master's thesis. California State University, Fresno.

Roberts, R. M. (2007, April). *The effective instructional use of computers: A meta-analysis*. Paper presented at the Annual meeting of the American Educational Research Association, Chicago, IL, April, 2007.

Robinson, D. H. & Schraw, G. (Eds.). (2008). A need for quality research in E-learning. In *Recent innovations in educational technology that facilitate student learning*. Charlotte, NC: Information Age Publishing.

Robles, M., & Braathen, S. (2002). Online assessment techniques. *The Delta Pi Epsilon Journal*, 44(1), 39-49.

Roblyer, M. D. (2005). Educational technology research that makes a difference: Series introduction. *Contemporary Issues in Technology and Teacher Education [Online serial], 5*(2). http://www.citejournal.org/vol5/iss2/seminal/article1.cfm

Roblyer, M. D. (2006). A deconstructed example of a type 2 study: Research to improve implementation strategies. *Contemporary Issues in Technology and Teacher Education* [Online serial], *6*(3). Available: http://www.citejournal.org/vol6/iss3/seminal/article1.cfm

Roblyer, M. D. (2007). A deconstructed example of a Type 4 study: Research to monitor and report on common uses and shape desired directions. Article submitted to *Contemporary Issues in Technology and Teacher Education (CITE) Journal*.

Romiszowski, A. J. (1981). *Designing instructional systems*. London: Kogan Page.

* Romanov, K. & Nevgi, A. (2006). Learning outcomes in medical informatics: Comparison of a WebCT course with ordinary web site learning material. *International Journal of Medical Informatics, 75*, 156—162. doi:10.1016/j.childyouth.2006.02.003

Rosen, Y. & Salomon, G. (2007). The differential learning achievements of constructivist technology-intensive learning environments as compared with traditional ones: a meta-analysis. *Journal of Educational Computing Research, 36*(1), 1-14.

Rosenthal, M. C., & Rubin, D.B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74,* 166-169.

Rosenthal, R. (1979). The file-drawer problem and tolerance for null results. *Psychological Bulletin*, *86*, 638-641.

Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.

Rosenthal, R. (1990). An evaluation of procedure and results. In K. W. Wachter & M. L. Straf, (Eds.), *The future of meta-analysis* (pp. 123-133). New York: Russell Sage Foundation.

Rowland, F., & Volet, S. (1996, July). Self-direction in community learning: a case study. *Australian Journal of Adult and Community Education*, 36(2), 89-102. (EJ 535 209)

* Ruksasuk, N. (2000). Effects of learning styles and participatory interaction modes on achievement of Thai students involved in web-based instruction in library and information science distance education. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9974471).

Russell, T. L. (1999). *The "no significant difference phenomenon."* Retrieved September 25, 1998 from http://teleeducation.nb.ca/nosignificantdifference/

Sabelli, N. (2004). Policy, planning, and the evaluation of learning technology. In Means, B. & Haertel, G. (Eds.), *Using technology evaluation to enhance student learning.* New York: Teachers College Press.

Salomon, G. & Perkins, D. (2005). Do technologies make us smarter? Intellectual amplification with, of and through technology. In Sternberg, R. J. & Priess, D. D. (Eds.), *Intelligence and technology: The impact of tools on the nature and development of human abilities,* (pp. 71-86). Mahweh, NJ: Erlbaum and Associates.

Schwab, J. H. (1983). The practical 4: something for curriculum professors to do. *Curriculum Inquiry, 13*(3), 239-265.

Schwarzer, R. (1991). Statistics software for meta-analysis (Version 5.3). [Computer software and Manual]. Berlin: Frie Universität Berlin. Retrieved from http://www.yorku.ca/faculty/academic/schwarze/meta_e.htm

* Seabolt, B. E. (2008). *A comparison of two delivery formats of ESL professional development*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3318639).

* Sendag, S. & Odabasi, H. F. (2009). Effects of an online problem based learning course on content knowledge acquisition and critical thinking skills. *Computers & Education, 53*, 132–141. doi:10.1016/j.compedu.2009.01.008

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261–281). New York: Russell Sage Foundation.

Shadish, Jr. W. R. & Sweeney, R. B. (1991). Mediators and moderators in meta-analysis: There's a reason we don't let dodo birds tell us which psychotherapies should have prizes. *Journal of Counseling and Clinical Psychology, 59*(6), 883-893.

* Shana, Z. (2009). Learning with Technology: Using discussion forums to augment a traditional-style class. *Educational Technology & Society*, *12* (3), 214–228.

Shield, G. (2000). A critical appraisal of learning technology using information and communication technologies. *Journal of Technology Studies*, *26*(1). [online] Available at: http://scholar.lib.vt.edu/ejournals/JOTS/Winter-Spring-2000/pdf/shield.pdf

Shulman, L. S. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher, 15*(X), 4-14.

Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2005). The comparative effectiveness of web-based and classroom instruction: A meta-analysis. *Personnel Psychology*, 2006. Available at http://interscience.wiley.com

Skinner, R. (2007). *Individual learning style preferences: A generational comparison in beginning education classes in selected Arizona community colleges*. (Doctoral dissertation, University of Nebraska-Lincoln, 2007). Retrieved from ProQuest database January 28, 2011. (UMI No. 3252841)

* Skylar, A. A. (2004). *Distance education: An exploration of alternative methods and types of instructional media in teacher education*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3144530)

Smith, M. C., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.

Song, F., Khan, K. S., Dinnes, J. & Sutton, A. J. (2002). Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *International Journal of Epidemiology, 31*, 88-95.

* Stanley, O. L. (2006). A comparison of learning outcomes by 'in-course' evaluation techniques for an on-line course in a controlled environment. *The Journal of Educators Online, 3*(2), 1-16. [Online]

Sterne, J. A.C. & Egger, M. (2001, October). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology, 54*(10), 1046-1055.

Tallent-Runnels, M. K., Thomas, J. A., Lan, W.Y, Cooper, S, Ahern, T. C., Shaw, S. M., & Liu, X. (2006). Teaching courses online: A review of the research. *Review of Educational Research, 76*(1), pp. 93-135.

Tam, M. (2009). Contructivism, instructional design, and technology: Implications for transforming distance learning (pp. 61-80). In Willis, J. B. (Ed.), *Constructivist instructional design (C-ID): foundations, models and examples*. Charlotte, NC: Information Age Publishing, Inc.

Thirunarayanan, M. O., & Perez-Prado, A. (2001). Comparing web-based and class-room based learning: a quantitative study. *Journal of Research on Technology in Education, 34*, 131-137.

Thornburg, D. D. (1996). *Campfires in cyberspace*. San Carlos, CA: Starsong Publications.

* Tsai, P., Hang, G., Tseng, J. C. R., & Hwang, G. (2008). A computer-assisted approach to conducting cooperative learning process. *International Journal of Distance Education Technologies, 6*(1), 49-66.

Tynjälä, R. (1999) Towards expert knowledge? A comparison between a constructivist and a traditional learning environment in the university, *International Journal of Educational Research*, 31, 357–442.

Uebersax, J. S. (1987). Diversity of decision making models and the measurement of interrater agreement. *Psychological Bulletin*, *101*, 140–146.

Ulmer, G. (1998). Foreword/Forward (Into Electracy) in T. Taylor, & I. Ward (Eds.), *Literacy theory in the age of the Internet*. New York: Columbia University Press.

Ungerleider, C. & Burns, T. (2003). *A systematic review of the effectiveness and efficiency of networked ICT in education. A State of the Field Report to the Council of Ministers of Education, Canada and Industry Canada*. Retrieved from http://www.cmec.ca/stats/SystematicReview2003.en.pdf 1/27/2010

U.S. Department of Education. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Washington, D.C.: Office of Planning, Evaluation, and Policy Development. Retrieved from http://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.doc

U.S. Department of Education, Office of Educational Research and Improvement. (1994). *A preliminary study of the feasibility and utility for national policy of instructional good practice' indicators in undergraduate education*. ERIC Document Reproduction Service No. ED 372 718. Washington, DC: U.S. Government Printing Office.

United States Distance Learning Association, The. (USDLA). (2007). *Research, statistics and distance learning resourced*. Retrieved July 2, 2007, from http://www.usdla.org/html/aboutUs/researchInfo.htm

Vines, D. L. & Rowland, J. R. (1995). An instructional feedback model for improved learning and mentoring. In *Proceedings of the Frontiers in Education Conference, Atlanta, GA, Nov. 1995*. DOI: 10.1109/FIE.1995.483145

Vonderwell, S. (2003). An examination of asynchronous communication experiences and perspectives of students in an online course: a case study. *Internet and Higher Education, 6*(1), 77-90.

* Wallace, T. Grinnell, L. Carey, L. & Carey, J. (2006). Maximizing learning from rehearsal activity in Web-based distance learning. *Journal of Interactive Learning or exercise and Research, 17*(3), 319-327.

West, C. R., Kahn, J. H., & Nauta, M. M. (2007). Learning styles as predictors of self-efficacy and interest in research: Implications for graduate research training. *Training and Education in Professional Psychology, 1*(3), 174-183.

WestEd. (2001). *Making time count*. Policy brief. San Francisco: Author. Available at: http://web.wested.org/online_pubs/making_time_count.pdf

* Westhius, D. Ouellette, P. M. & Pfahler, C. L. (2006). A comparative analysis of on-line and classroom-based instructional formats for teaching social work research. *Advances in Social Work, 7*(2), 74-88.

Wiburg, K. (2009). Instructional design: Is it time to exchange Skinner's teaching machine for Dewey's toolbox? (pp. 47-60). In Willis, J. B. (Ed.), *Constructivist instructional design (C-ID): foundations, models and examples*. Charlotte, NC: Information Age Publishing, Inc.

* Williams, P. B. (2005). *On-demand tutoring in distance education: Intrinsically-motivated, scalable interpersonal interaction to improve achievement, completion, and satisfaction*. (Doctoral dissertation). Available from ProQuest\ Dissertations and Theses database. (UMI No. 3194547)

Wilson, D. B., & Lipsey, M. W. (2001 B). The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods, 6*(4), 413-429.

Winograd, D. (2000, October). *The effects of trained moderation in online asynchronous distance learning*. Paper presented at the annual meeting of the Association for Educational Communication and Technology, Denver, Co.

* Wise, A., Chang, J., Duffy, T., & Del Valle, R. (2004). The effects of teacher social presence on student satisfaction, engagement, and learning. *Journal of Educational Computing Research, 31*(3) 247-271.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: Sage.

Wolters, C. (1998). Self-regulated learning and college students' regulation of motivation. *Journal of Educational Psychology, 90*(1), 224-235.

* Yang, Y. C., Newby, T. & Bill, R. (2008). Facilitating interactions through structured web-based bulletin boards: A quasi-experimental study on promoting learners' critical thinking skills. *Computers & Education, 50*, 1572–1585. doi:10.1016/j.compedu.2007.04.006

* Yavuz, S. (2007, Jan.). The effects of interactive learning environments on cooperative learning achievement and student anxiety in environmental education. *Turkish Online Journal of Distance Education-TOJDE, 9*(1), Article 13.

Young, S. S. (2004). In search of online pedagogical models: Investigating a paradigm change in teaching through the School for All community. *Journal of Computer Assisted Learning, 20*, 133-150.

Yuen, K. & Hau, K. (2006). Constructivist teaching and teacher-centred teaching: a comparison of students' learning in a university course. *Innovation in Education and Teaching International, 43*(3), 279-290. doi: 10.1080/14703290600750861

Zhang, D. Zhou, L. Briggs, R. O. & Nunamaker, Jr., J. F. (2006). Instructional video in e-learning: assessing the impact of interactive video on learning effectiveness. *Information and Management*, *43*(1), 15-27.

Zhang, L. J. (2001). Awareness in reading: EFL students' metacognitive knowledge of reading strategies in an input-poor environment. *Language Awareness*, *11*(4), 268-288.

Zhao, Y., Lei, J., Yan, B., Lai, C. & Tan, H. S. (2005, August). What makes the difference? A practical analysis of research on the effectiveness of distance education. *Teachers College Record*, *107*(8), 1836-1884.

Zimmerman, B. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, *81*(3), 329-339.

VITA

Graduate College
University of Nevada, Las Vegas

Robin Michael Roberts

Degrees:
Bachelor of Arts, Industrial Arts and Science, 1978
California State University, Fresno

Master of Arts, Christian Education, 1981
Mennonite Brethren Biblical Seminary, Fresno, CA

Certificate of Advanced Studies, Instructional Technology, 2001
California State University, Fresno

Master of Arts, Education, 2002
California State University, Fresno

Special Honors and Awards:
Graduate Dean's Medal, Kreman School of Education, 2003
California State University, Fresno

Outstanding Thesis Award Nominee, Kreman School of Education, 2003
California State University, Fresno

Dee Smith Scholarship (Graduate), 2006
University of Nevada, Las Vegas

UNLV Summer Session Scholarship (Graduate), 2007
University of Nevada, Las Vegas

Publications:
Books

Roberts, R. M. (2008). *Images of America: Kings County*. Charleston, SC: Arcadia Publishing Co. ISBN 0-7385-5844-3

Roberts, R. M. (2007). *Images of America: Hanford, 1900 – 2000*. Charleston, SC: Arcadia Publishing Co. ISBN 0-7385-4768-9

Roberts, R. M. (2005). *Images of America: Hanford, California*. Charleston, SC: Arcadia Publishing Co. ISBN 0-7385-2948-6

Chapters in Books

Roberts, R. M. (2009).The Digital Generation and Web 2.0: E-Learning Concern or Media Myth? In H. Yang & S. Yeun (Eds.), *Handbook of Research on Practices and Outcomes in E-Learning: Issues and Trends,* Chapter 6. Hershey, PA: IGI Global.

Roberts, R. M., (2009). General nature of the survey area. In Russell, E., Arroues, K. D. & Schmidt, B., *Soil survey of Kern county, California, southwestern part*. USDA Natural Resources Conservation Service. Washington, D.C.: GPO. Available at: http://soils.usda.gov/survey/online_surveys/california/#kern2009

Roberts, R. M. (2007). General nature of the survey area. In Arroues, K.D., Russell, E. & Regal, J., *Soil survey of Kern county, northeastern part, and southeastern part of Tulare County, California*. USDA Natural Resources Conservation Service. Washington, D.C.: GPO. Available at: http://soils.usda.gov/survey/online surveys/california/#kern2007

Roberts, R. M. (2006). General nature of the survey area. In Arroues, K. *Soil survey of Fresno county, California, western part*. USDA Natural Resources Conservation Service. Washington, D.C.: GPO.

Presentations, Conference Papers, Seminars and Workshops (selected)

Schraw, G., Kuch, F. & **Roberts, R.** (2011, April). *Bias in the gamma coefficient: a Monte Carlo study*. Paper presented for the Symposium "Calibrating Calibration: Conceptualization, measurement, calculation, and context" at the Annual meeting of the American Educational Research Association, New Orleans, LA., April 10, 2011.

Kuch, F., Schraw, G., & **Roberts, R.** (2010, October). *Bias in the gamma coefficient: A Monte Carlo study.* Paper presented at the 28th Annual NRMERA Conference, Big Sky, Montana, October 1, 2010.

Roberts, R. M. & Strudler, N. (2009, April). *The differential use of Web 2.0 technologies by Digital Generation pre-service teachers*. Paper presented at the Annual meeting of the American Educational Research Association, San Diego, CA, April 14, 2009.

Roberts, R. M. & Strudler, N. (2008, March). *Some attributes of 21st Century learners in the online environment: The Digital Generation and Web 2.0*. Paper presented at SITE 2008--Society for Information Technology & Teacher Education International Conference, Las Vegas, Nevada, March 7, 2008.

Roberts, R. M. (2007, April). *The effective instructional use of computers: A meta-analysis*. Paper presented at the Annual meeting of the American Educational Research Association, Chicago, IL, April 2007.

Dissertation Title: Best Instructional Practices for Distance Education: A Comparative Meta-Analysis

Dissertation Examination Committee:
Chairperson, Dr. Neal B. Strudler, Ph.D.
Committee Member, Dr. Kendall Hartley, Ph.D.
Committee Member, Dr. Gregory Schraw, Ph.D.
Graduate Faculty Representative: Dr. James R. Crawford, Ph.D.