

Running Head: Meta-Analysis of Distance Education Studies

How Does Distance Education Compare to Classroom Instruction?

A Meta-Analysis of the Empirical Literature

Robert M. Bernard, Philip C. Abrami, Yiping Lou¹, Evgueni Borokhovski, Anne Wade,
Lori Wozney, Peter Andrew Wallet, Manon Fiset and Binru Huang¹

Centre for the Study of Learning and Performance
Concordia University, Montreal, QC, Canada

¹Louisiana State University, Baton Rouge, LA, U.S.A.

Review of Educational Research (in press)

<<PLEASE DO NOT QUOTE OR CITE WITHOUT PERMISSION>>

June 22, 2004

This study was supported by grants from the *Fonds québécois de la recherche sur la société et la culture* and the Social Sciences and Humanities Research Council of Canada to Abrami and Bernard, and funding from the Louisiana State University Council on Research Awards and Department to Lou. The authors express appreciation to Janette M. Barrington, Anna Peretiatkovicz, Mike Surkes, Lucie A. Ranger, Claire Feng, Vanikumari Pulendra, Keisha Smith, Alvin Gautreaux, and Venkatraman Kandaswamy for their assistance and contributions. The authors thank Dr. Richard E. Clark, Dr. Gary Morrison, and Dr. Tom Cobb for their comments on this research and their contributions to the development of ideas for analysis and discussion. Contact: Dr. Robert M. Bernard, CSLP, LB-545-5, Department of Education, Concordia University, 1455 de Maisonneuve Blvd. W., Montreal, QC, Canada H3G 1M8. E-mail: bernard@education.concordia.ca Website: <http://doe.concordia.ca/cslp>

Abstract

A meta-analysis of the comparative distance education (DE) literature between 1985 and 2002 was conducted. In total, 232 studies containing 599 independent achievement, attitude, and retention outcomes were analyzed. Overall results indicated effect sizes of essentially zero on all three measures and wide variability. This suggests that many applications of DE outperform their classroom counterparts and many applications perform more poorly. Dividing achievement outcomes into synchronous and asynchronous forms of DE produced a somewhat different impression. In general, mean achievement effect sizes for synchronous applications favored classroom instruction while for asynchronous applications they favored DE. However, significant heterogeneity remained in each subset. Three clusters of study features—research methodology, pedagogy, and media—entered into weighted multiple regression, revealed, in general, that methodology accounted for the most variation followed by pedagogy and media, suggesting that Clark's (1983, 1994) claims of the importance of pedagogy over media are essentially correct. We go on to suggest that researchers move beyond simple comparisons between DE and classroom instruction to more pressing and productive lines of inquiry that may contribute more to our knowledge of what works best in DE.

Introduction

In the same way that transitions between technological epochs often breed transitional names that are shed as the new technology becomes established (e.g., the automobile was called the “horseless carriage” and the railroad train was called an “iron horse”), research in new applications of technology in education has initially focused on comparisons with more established instructional applications, such as classroom instruction. In the 1950s and 60s, the emergence of television as a new medium of instruction initiated a flurry of research that compared it with “traditional” classroom instruction. Similarly, various forms of computer-based instruction (1970s and 80s), multi-media (1980s and 90s), teleconferencing (1990s), and distance education (spanning all of these decades) have been investigated from a comparative perspective in an attempt to judge their relative effectiveness. It is arguably the case that these comparisons are necessary for policymakers, designers, researchers, and adopters to be certain of the relative value of innovation. Questions about relative effectiveness are important both in the early stages of development and as a field matures to summarize the nature and extent of the impact on important outcomes, giving credibility to change and helping to focus it.

This study deals specifically with comparative studies of distance education. Keegan’s (1996) definition of distance education (DE) is perhaps the most commonly cited in the literature, and involves five qualities that distinguish it from other forms of instruction: a) the quasi-permanent separation of teacher and learner; b) the influence of an educational organization, both in planning, preparation, and the provision of student support; c) the use of technical media; d) the provision of two-way communication; and e) the quasi-permanent absence of learning groups. This latter element has been debated in the literature (Garrison & Shale, 1987; Verduin & Clark, 1991) because it seemingly excludes many applications of DE based upon teleconferencing technologies that are group-based. Some argue that when DE simply recreates the conditions of a traditional classroom it misses the point because DE of this type does not support the “anytime, anyplace” objective of access to education for students who cannot be in a particular place at a particular time. However, synchronous DE does fall within the purview of current practices and therefore qualifies for consideration. To Keegan’s definition, Rekkedal and Qvist-Eriksen (2003) add the following adjustments to accommodate e-learning:

- the use of *computers and computer networks* to unite teacher and learners and carry the content of the course;
- the provision of two-way communication *via computer networks* so that the student may benefit from or even initiate dialogue (this distinguishes it from other uses of technology in education). (p. 1)

In characterizing DE, Keegan also distinguishes between “distance teaching” and “distance learning.” It is a fair distinction that applies to all organized educational events. Since learning does not always follow from teaching, it is also a useful way of discussing the elements—teaching and learning—that constitute a total educational setting. While Keegan does not go on to explain, specifically, how these differ in practice, it can be assumed that teaching designates activities in which teachers engage (e.g., lecturing, questioning, providing feedback), while learning designates activities in which students engage (e.g., taking notes, studying, reviewing, revising).

The media used in DE have undergone remarkable changes over the years. Taylor (2001) characterizes five generations of DE, largely defined with regard to the media, and thereby the range of instructional options available at the time of their prevalence. The progression that

Taylor describes moves along a rough continuum of increased flexibility, interactivity, materials delivery, and access, beginning in the early years of DE when it was called correspondence education (i.e., the media were print and the post office), through broadcast radio and television, and on to current manifestations of interactive multi-media, the Internet, access to Web-based resources, computer-mediated communication, and, most recently, campus portals providing access to the complete range of university services and facilities at a distance. Across the history of DE research, most of these media have been implicated in DE studies in which comparisons have been made to what is often referred to as “traditional classroom-based instruction” or “face-to-face” instruction. It is this literature that is the focus of this meta-analysis.

Instruction, Media, and DE Comparison Studies

Clark (1983, 1994) rightly criticized early media comparison studies on a variety of grounds, the most important of which is that the medium under investigation, the instructional method that is inextricably tied to it, and the content of instruction, together, form a confound that renders their relative contributions to achieving instructional goals impossible to untangle. Clark goes on to argue that the instructional method is the “active ingredient,” not the medium—the medium is simply a neutral carrier of content and of method. In essence, he argues that any medium, appropriately applied, can fulfill the conditions for quality instruction, and so cost and access should form the decision criteria for media selection. Effectively, these arguments suggest that media serve a transparent purpose in DE.

Several notable rebuttals of Clark’s position have followed (Ullmer, 1994; Kozma, 1994; Morrison, 1994; Tennyson, 1994). Kozma argued that Clark’s original assessment was based on “old non-interactive technologies” that simply carried method and content, where a distinction between these elements could be clearly drawn. More recent media uses, he added, involve highly interactive sets of events that occur between learners and teachers, among learners (e.g., collaborative learning), often within a constructivist framework, and even between learners and non-human agents or tools, so that a consideration of discrete variables no longer makes sense. The distinction here seems to be “media to support teaching” and “media to support learning,” which is completely in line with Keegan’s reference to distance teaching and distance learning.

Cobb (1997) added an interesting wrinkle to the debate. He argued that under certain circumstances, the efficiency of a medium or symbol system can be judged by how much of the learner’s cognitive work it performs. By this logic, some media, then, have advantages over other media, since it is “easier” to learn some things with certain media than with others. The way to advance media design, according to Cobb, “is to model learner and medium as distributed information systems, with principled, empirically determined distributions of information storage and processing over the course of learning” (p. 33). According to this argument, the medium becomes the tool of the learner’s cognitive engagement and not simply an independent and neutral means for delivering content. It is what the learner does with a medium that counts, not so much what the teacher does. These arguments suggest that media are more than just transparent, they are also transformative.

Why Do Comparative DE Studies?

One of the differences between DE and media comparison studies is that DE is not a medium of instruction, but it depends entirely on the availability of media for delivery and communication (Keegan, 1996). DE can be non-interactive or highly interactive and may, in fact, encompass one or many media types (e.g., print + video + computer-based simulations + computer conferencing) in the service of a wide range of instructional objectives. In the same way, classroom instruction may include a wide mix of media forms. So, in a well-conceived and

Bernard et al.

executed comparative study, where all of these aspects are present in both conditions, differences may relate more to the proximity of learner and teacher, one of Keegan's defining characteristics of DE, and differential means through which interaction and learner engagement can occur. Synchronicity and asynchronicity, as well as the attendant issues of instructional design, student motivation, feedback and encouragement, direct and timely communication, and perceptions of isolation might then form the major distinguishing features of DE and classroom instruction. Shale (1990) comments "In sum, DE ought to be regarded as education at a distance. All of what constitutes the process of education when teacher and student are able to meet face-to-face also constitutes the process of education when teacher and student are physically separated." (p. 334)

This, in turn, suggests that "good" DE applications and "good" classroom instruction should be, in principle, relatively equal to one another, regardless of the media used, especially if a medium is used simply for the delivery of content. However, when the medium is placed in the hands of learners to make learning more constructive or more efficient, as suggested by Kozma and Cobb, the balance of effect may shift. In fact, in DE, media may transform the learning experience in ways that are unanticipated and not regularly available in face-to-face instructional situations. For example, the use of computer-mediated communication means students must use written forms of expression to interact with one another in articulating and developing ideas, arguing contrasting viewpoints, refining opinions, settling disputes and so on (Abrami & Bures, 1996). This use of written language and peer interaction may result in increased reflection (Hawkes, 2001) and the development of better writing skills (Winkelmann, 1995). Higher quality performance of complex problem-solving skills may develop through peer modeling and mentoring (Lou, 2004; Lou, Deidic & Rosenfield, 2003; Lou & MacGregor, 2002). The critical thinking literature goes so far as to suggest that activity of this sort can promote the development of critical thinking skills (Garrison, Anderson & Archer, 2001; McKnight, 2001).

Is it necessary or even desirable, then, to continue to conduct studies that directly compare DE with classroom teaching? Clark (2000), by exclusion, claims that it is not: "... all evaluations should explicitly investigate the relative benefits of two different but compatible types of DE technologies found in every DE program." (p. 4). By contrast, Smith and Dillon (1999) argue that comparative studies are still useful, but only when they are done in light of a full analysis of media attributes and their hypothesized effects on learning, and when these same attributes are present and clearly articulated in the comparison conditions. In the eyes of Smith and Dillon, it is only under these circumstances that comparative studies can push forward our understanding of the features of DE and classroom instruction that make them similar and/or different. Unfortunately, as Smith and Dillon point out, this level of analysis and clear accounting of the similarities and differences between treatment and control is not often reported in the literature, and so it is difficult to determine the existence of confounds across treatments which would render such studies uninterpretable.

There may be a more practical reason for assessing the effectiveness of DE in comparison to its classroom alternatives. There was a time when DE was regarded simply as a reasonable alternative to campus-based education, primarily for students who had restricted access to campuses because of geography, time constraints, disabilities, or other circumstances. And by virtue of the limitations of the communication facilities that existed at that time (e.g., mail, telephone, TV coverage), DE itself tended to be restricted by geographical boundaries (e.g., for many years the UK Open University was available only to students in Britain). However, the reality of "learn anywhere, anytime," promulgated largely by the communication and technological resources offered by the Internet and broadband ISPs, have set traditional educational institutions into intense competition for the world-wide market of "online learners." So it is arguable that finding answers to the question that has guided much of the comparative

Bernard et al.

research on DE in the past—Is distance learning as effective as classroom learning? —has become even more pressing. Should educational institutions continue to develop and market Internet learning opportunities without knowing whether they will be as effective as their classroom-based equivalents or, in the worse case, whether they will be effective at all? Based on longstanding instructional design thinking, it is not enough to develop a technology-based course simply because the technology of delivery exists, and yet the reverse of this very thinking seems to prevail in the rush to get courses and even whole degree programs online. Beyond being simply a “proof of worthiness,” well-designed studies can suggest to administrators and policymakers not only *whether* DE is a worthwhile alternative, but also in which content domains, with which learners, under what pedagogical circumstances, and with which mix of media the transformation of courses and programs to DE is justified. In fact, it is not unreasonable to suggest that such studies might be conducted under “local circumstances” for the primary purpose of making decisions that affect institutional growth on that particular campus.

Evidence of Effectiveness

The answer to the DE effectiveness question, or any research question for that matter, cannot be found in a single study. It is only through careful reviews of the general state of affairs in a research literature that large questions can be addressed, and that the quality of the research itself and the veracity of its findings can be assessed.

There have been many attempts to summarize the comparative research literature of DE. The most comprehensive, but least assiduous, is Russell’s (1999) collection of 355 “no significant difference” studies. Based on evidence in the form of fragmented annotations (e.g., ... no significant difference was found ...) of all of the studies that could be located and contrasting this with the much smaller number of “significant difference studies” (which can be either positive or negative), Russell declared that there is no compelling evidence to refute Clark’s original 1983 claim that a delivery medium contributes little if anything to the outcomes of planned instruction, and, by extension, that there is no advantage in favor of technology-delivered DE. But there are several problems with Russell’s approach. First, not all studies are of equal quality and rigor, and to include them all, without qualification or description, renders conclusions and generalizations suspicious, at best. Second, an accepted null hypothesis does not deny the possibility that unsampled differences exist in the population; it only means that they do not exist in the sample being studied. This is particularly true in small-sample studies where power to reject the null hypothesis (and thus the risk of making Type II errors) is high. Third, differential sample sizes of individual studies make it impossible to aggregate the results of different studies solely on the basis of their test statistics. So Russell’s work represents neither a sufficient overall test of the hypothesis of no difference nor an estimate of the magnitude of effects attributable to DE.

Another widely cited report (Phipps & Merisotis, 1999), prepared for the American Federation of Teachers and the National Education Association, and entitled “What’s the Difference: A Review of Contemporary Research on the Effectiveness of Distance Education in Higher Education,” may contain a similar level of bias to that in Russell’s work, but for a different reason. In the words of the authors, “While this review of original research does not encompass every study published since 1990, it does capture the most important and salient of these works.” (p. 154). In fact, just over 40 empirical investigations are cited to illustrate specific points that are made by the authors. The problem is, how can we judge importance or salience without carefully crafted inclusion and exclusion criteria? The bias that is risked, then, is one of selecting research, even unconsciously, to make a point, rather than accurately characterizing the state of the research literature around a given question. While one of the findings of the report

Bernard et al.

may generally be true—that the literature lacks rigor of methodology and reporting—the findings of the “questionable effectiveness of DE” based on a select number of studies is no more credible than Russell’s claim of non-significance based on everything that has ever been published. Somewhere between these extremes lies evidence that can be taken as more representative of the true state of affairs in the population.

In addition to these reports, there have been a number of more or less extensive narrative reviews of research (e.g., Berge & Mrozowski, 2001; Saba, 2000; Jung & Rha, 2000; Schlosser & Anderson, 1994; Moore & Thompson, 1990). This type of research has long been known for its subjectivity, potential bias, and inability to answer questions about magnitude of effects.

Meta-analysis or quantitative synthesis, developed by Gene Glass and his associates (Glass, McGaw & Smith, 1981), represents an alternative to the selectivity of narrative reviews and the problem of conclusions based on test statistics from studies with different sample sizes. Meta-analysis makes it possible to combine studies with different sample sizes by extracting an effect size from all studies. Cohen’s *d* is a sample-size-based index of standardized differences between a treatment and control group, which can be averaged in a way that test statistics cannot. Refinements by Hedges and Olkins (1985) further reduce the bias resulting from differential sample sizes among studies. Thus a meta-analysis is an approach to estimating how much one treatment differs from another, over a large set of similar studies, and the variability that is associated with it. An additional advantage of meta-analysis is that moderator variables can be investigated to explore more detailed relationships that may exist in the data.

A careful analysis of the accumulated evidence of DE studies can allow us to estimate mean effect size and variability in the population and to explore what might be responsible for variability in findings across media, instructional design, course features, students, settings, etc. Research methodology can also be investigated, therefore shedding light on some of the issues of media, method, and experimental confounds pointed out by Clark and others. At the same time, failure to reach closure on these issues exposes the limitations in the existing research base, both in quantity and quality, indicating directions for further inquiry.

In summary, meta-analysis has the following advantages: a) it answers questions about size of effect; b) it allows systematic exploration of sources of variability in effect size; c) it allows for control over *internal validity* by focusing on comparison studies vs. one-shot case studies; d) it maximizes *external validity* or generalizability by addressing a large collection of studies; e) it improves statistical power when a large number of studies is analyzed; f) it uses the student as the unit of analysis, not the study—large sample studies have higher weight; g) it allows new studies to be added as they become available or studies to be deleted as they are judged to be anomalous; h) it allows new study features and outcomes to be added to future analyses as new directions in primary research emerge; i) it allows analysis and re-analysis of parts of the dataset for special purposes (e.g., military studies, synchronous vs. asynchronous instruction, web-based instruction); and j) it allows comment on what we know and what we need to know (Bernard & Naidu, 1990, Abrami, Cohen & d’Apollonia, 1988).

Five quantitative syntheses that are specifically related to DE and its correlates have been published (Shachar & Neumann, 2003; Ungerleider & Burns, 2003; Allen, Bourhis, Burrell & Mabry, 2002; Cavanaugh, 2001; Machtmes & Asher, 2000). In the most recent meta-analysis, Shachar and Neumann reviewed 86 studies, dated between 1990 and 2002, and found an effect size for student achievement of +0.37, which, if it holds up, belies the general impression given by other studies that DE and classroom instruction are relatively equal. In another recent study, Ungerleider and Burns did a systematic review for the Council of Ministers of Education, Canada (CMEC), including a quantitative meta-analysis of the literature of networked and online learning (i.e., not specifically DE). They found poor methodological quality, to the extent that

only 12 achievement and four satisfaction outcomes were analyzed. They also found an overall effect size of zero for achievement and an effect size of -0.509 for satisfaction. Both findings were significantly heterogeneous. Here is an example of two credible works providing conflicting evidence as to the state of comparative studies.

Allen, Bourhis, Burrell, and Mabry summarized 25 empirical studies in which DE and classroom conditions were compared on the basis of measures of student satisfaction. Studies were excluded from consideration if they did not contain a comparison group and did not report sufficient statistical information from which effect sizes could be calculated. The results revealed a slight correlation ($r = +0.031$, $k = 25$, $N = 4,702$, significantly heterogeneous sample) favoring classroom instruction. When three outliers were removed from the analysis, the correlation coefficient increased to 0.090 , and the homogeneity assumption was satisfied. Virtually no effects were found for “channel of communication” (video, audio, and written) or its interaction with “availability of interaction.” This meta-analysis is limited in that it investigates only one outcome measure, student satisfaction, arguably one of the least important indicators of effectiveness, and its sample size and range of coded moderator variables yield little more than basic information related to the question of DE effectiveness.

The Cavanaugh meta-analysis examined interactive (i.e., videoconferencing and telecommunications) DE technologies in K–12 learning in 19 experimental and quasi-experimental studies on the basis of student achievement. Studies were selected on the following bases: a) they included a focus on interactive DE technology; b) they were published between 1980 and 1998; c) they included quantitative outcomes from which effect sizes could be extracted; and d) they were free from obvious methodological flaws. In 19 studies ($N = 929$) that met these criteria, results indicated an overall effect size (i.e., weighted mean difference) of $+0.015$ in favor of DE conditions for a significantly heterogeneous sample. This effect size was considered to be *not significant*. Subsequent investigation of moderator variables revealed no additional findings of consequence. This study is limited in its purview to K-12 courses, generalizing to what is perhaps the least developed “market” for DE.

The fourth meta-analysis performed by Machtmes and Asher compared live or pre-produced adult tele-courses with their classroom equivalents on measures of classroom achievement in either experimental or quasi-experimental designs. Out of 30 studies identified, 19 studies dated between 1943 and 1997 were coded for effect size and study features. The overall weighted effect size for these comparisons was -0.0093 (not significant; ranging from $+1.50$ to -0.005). The assumption of homogeneity of effect size was violated, and was attributed to differences in learners’ levels of education and differences in technology over the period of time under consideration. Three study features were found to affect student achievement: type of interaction available, type of course, and type of remote site.

In the literature of DE comparison reviews, we find only fragmented and partial attempts to address the myriad of questions that might be answerable from the primary literature, but we also find great variability among findings but general agreement concerning the poor quality of the literature. In this era of proliferation of various technology-mediated forms of DE, it is time for a comprehensive review of the empirical literature to assess the quality of the DE research literature systematically, to attempt to answer questions relating to the effectiveness of DE, and to suggest directions for future practice and research.

Synchronous and Asynchronous DE

In the age of the Internet and CMC, there is a tendency to think of DE in terms of “anywhere, anytime education.” DE of this type truly fits two of Keegan’s (1996) definitional criteria, “the quasi-permanent separation of teacher and learner” and “the quasi-permanent

Bernard et al.

absence of learning groups.” However, much of what is called DE does not fit either of these two criteria, rendering it DE that is group-based and time- and place-dependent. This form of DE, which we will call synchronous DE, is not so very different from early applications of distributed education via closed-circuit television on university campuses (e.g., Penn State) that began in the late 1940s. The primary purpose of this movement in the U.S. was to economize on teaching resources and subject matter expertise by distributing live lectures and, later, mediated questioning and discussion, to many “television classrooms” or remote sites across a university campus or other satellite locales. Many studies of this form of instruction produced “no significant difference” between the live classroom and the remote site (e.g., Carpenter & Greenhill, 1955; 1958).

The term distance education became attached to this form of instruction as the availability and reliability of videoconferencing and interactive television began to emerge in the mid-1980s. The premise, however, remains the same: two or more classes in different locations connected via some form of telecommunication technology, being directed by one or more teachers. According to Mottet (1998) and Ostendorf (1997), this form of “emulated traditional classroom instruction” is the fastest growing form of DE in U.S. universities and so it is important for us to know how it affects learners who are involved in it.

Contrasted with this “group-based” form of instruction is “individually based” DE where, students in remote locations work independently or in asynchronous groups, usually with the support of an instructor or tutor. We have called this asynchronous because DE students are not synchronized with classroom students and because communication is largely asynchronous by e-mail or through CMC software. Chat rooms and the like offer an element of synchronicity, of course, but this is usually an optional feature of the instructional setting. Asynchronous DE has its roots in correspondence education, where learners were truly independent, connected to an instructor or tutor by the postal system; communication was truly asynchronous because of postal delay. Because of the differences in synchronous and asynchronous DE just noted, we decided to examine these two patterns undivided and divided. In fact, this distinction formed a natural division around which the majority of the analyses revolved.

For some, the key definitional feature of DE is the physical separation of learners in space and time. For others the physical separation in space only is a sufficient condition for DE. In the former definition, asynchronous communication is the norm. In the latter definition, synchronous communication is the norm.

We take no position on which of these definitions is correct, but note that there are numerous instances in the literature where both synchronous and asynchronous forms of communication are available to the learner. We have included both types in our review in order to examine how synchronicity and asynchronicity affect learning. Where a choice in design exists, knowing the influence of these patterns may guide instructional design. Where there is no choice in design and students must learn asynchronously, separated in both space and time, the development of new instructional resources as alternative supports for student learning needs may be needed.

There are, of course, hybrids of these two, referred to by some as “distributed education” (e.g., Dede, 1996). We did not attempt to separate these mixed patterns from those in which students truly worked independently from one another or in synchronous groups. So within asynchronous studies there is an element of within-group synchronicity (i.e., DE students communicating, synchronously, among themselves), just as there is some asynchronicity within synchronous studies. However, this does not affect the defining characteristics of synchronicity and asynchronicity as they are described here.

Statement of the Problem

The overall intention of this meta-analysis is to provide an exhaustive quantitative synthesis of the comparative research literature of DE, from 1985 to the end of 2002, across all age groups, media types, instructional methods, and outcome measures. From this literature it seeks to answer the following questions:

1. Overall, is interactive DE as effective, in terms of student achievement, attitudes, and retention, as its classroom-based counterparts?
2. What is the nature and extent of the variability of the findings?
3. How do conditions of synchronicity and asynchronicity moderate the overall results?
4. What conditions contribute to more effective DE as compared to classroom instruction?
5. To what extent do media features and pedagogical features moderate the influences of DE on student learning?
6. What is the methodological state of the literature? and
7. What are important implications for practice and future directions for research?

Method

This meta-analysis is a quantitative synthesis of empirical studies since 1985 that compared the effects of DE and traditional classroom-based instruction on student achievement, attitude, and retention (i.e., opposite of drop-out). The year 1985 was chosen as a cut-off date since electronically mediated, interactive DE became widely available around that time. The procedures employed to conduct this quantitative synthesis are described below under the following subheadings: working definition of DE, inclusion/exclusion criteria, data sources and search strategies, outcomes of the searches, outcome measures and effect size extraction, study features coding, and data analysis.

Working Definition of DE

The working definition of DE builds on Nipper's (1989) model of "third-generation distance learning," as well as Keegan's (1996) synthesis of recent definitions. Linked historically to developments in technology, first generation DE refers to the early days of print-based correspondence study. Characterized by the establishment of the Open University in 1969, second generation DE refers to the period when print materials were integrated with broadcast TV and radio, audio and videocassettes, and increased student support. Third generation DE was heralded by the invention of Hypertext and the rise in the use of teleconferencing (i.e., audio and video). To this, Taylor (2001) adds the "fourth generation," characterized by flexible learning (e.g., CMC, Internet accessible courses) and the "fifth generation" (e.g., online interactive multimedia, Internet-based access to WWW resources). Generations three, four and five represent moves away from directed and non-interactive courses to those characterized by a high degree of learner control and two-way communication, as well as group-oriented processes and greater flexibility in learning. With new communication technologies in hand and renewed interest in the convergence of DE and traditional education, this is an appropriate time to review the research on third, fourth, and fifth generation DE. Our definition of DE for the inclusion of studies thus reads:

- The semi-permanent separation (place and/or time) of learner and instructor during planned learning events.

Bernard et al.

- The presence of planning and preparation of the learning materials, student support services, and the final recognition of course completion by an educational organization.
- The provision of two-way media to facilitate dialogue and interaction between students and the instructor and among students.

Inclusion/Exclusion Criteria

To be included in this meta-analysis, each study had to meet the following inclusion/exclusion criteria:

1. Involve an empirical comparison of DE as defined in this meta-analysis (including satellite/TV/radio broadcast + telephone/e-mail, e-mail-based correspondence, text-based correspondence + telephone, web/audio/video-based two-way telecommunication) with face-to-face classroom instruction (including lectures, seminars, tutorials, and laboratory sessions). Studies comparing DE with national standards or norms, rather than a control condition, were excluded.
2. Involve “distance from instructor” as a primary condition of the DE condition. DE with some face-to-face meetings (less than 50%) was included. However, studies where electronic media were used to supplement regular face-to-face classes with the teacher physically present were excluded.
3. Report measured outcomes for both experimental and control groups. Studies with insufficient data for effect size calculations (e.g., with means but no standard deviations or no inferential statistics or no sample size) were excluded.
4. Be publicly available or archived.
5. Include at least one achievement, attitude or retention, outcome measure.
6. Include an identifiable level of learner. All levels of learners from kindergarten to adults, whether informal schooling or professional training, were admissible.
7. Be published or presented no earlier than 1985 and no later than December of 2002.
8. Include outcome measures that were the same or comparable. If the study explicitly stated that different exams were used for the experimental and control groups, the study was excluded.
9. Include outcome measures that reflected individual courses rather than whole programs. Thus, programs composed of many different courses, where no opportunity existed to analyze conditions and the corresponding outcomes for individual treatments, were excluded.
10. Include only studies with an N of 2 or greater (i.e., enough to form a standard deviation).
11. Include only the published source, when data about a particular study were available from different sources (e.g., journal article and dissertation). Additional data from the other source were used only to make coding study features more detailed and accurate.

Data Sources and Search Strategies

The studies used in this meta-analysis were located through a comprehensive search of publicly available literature from 1985 through December of 2002. Electronic searches were performed on the following databases: *ABI/Inform*, *Compendex*, *Cambridge Scientific Abstracts*, *Canadian Research Index*, *Communication Abstracts*, *Digital Dissertations on ProQuest*, *Dissertation Abstracts*, *Education Abstracts*, *ERIC*, *PsycInfo*, and *Social SciSearch*. Web

Bernard et al.

searches were performed using Google, AlltheWeb, and Teoma search engines. A manual search was performed in *ComAbstracts*, *Educational Technology Abstracts*, and in several distance learning journals, including *The American Journal of Distance Education*, *Distance Education*, *Journal of Distance Education*, *Open Learning*, and *Journal of Telemedicine and Telecare*; and in several conference proceedings, including *ACE*, *AERA*, *CADE*, *EdMedia*, *E-Learn*, *SITE*, and *WebNet*. In addition, the reference lists of several earlier reviews including, Moore and Thompson (1990); Russell (1999); Machtmes and Asher (2000); Cavanaugh (2001); Allen, Bourhis, Burrell, and Mabry (2002); and Shachar (2002), were reviewed for possible inclusions. Although search strategies varied depending on the tool used, generally, search terms included “distance education,” “distance learning,” “open learning” or “virtual university,” AND (“traditional,” “lecture,” “face-to-face,” or “comparison”).

Outcomes of the Searches

In total, 2,262 research abstracts concerning DE and traditional classroom-based instruction were examined and 862 full-text potential includes retrieved. Each of the studies retrieved was read by two researchers for possible inclusion using the inclusion/exclusion criteria. The initial inter-rater agreement as to inclusion was 89%. Any study that was considered for exclusion by one researcher was crosschecked by another researcher. Two hundred and thirty-two (232) studies met all inclusion criteria and were included in this meta-analysis; 630 were excluded. Table 1 shows the categories of exclusion and the number and percentage of the excluded studies.

Table 1
Category, Number and Percentage of Excluded Studies

Category	Excluded Studies	
	<i>N</i>	%
1. Review and conceptual articles	52	8.25
2. Case studies, survey results and qualitative studies	55	8.73
3. Studies with some violation of either DE or F2F definitions	295	46.83
4. Collapsed data, mixed conditions or program-based findings	43	6.83
5. Insufficient statistical data	97	15.40
6. Non-retrievable studies	10	1.58
7. “Out-of-date” studies	21	3.33
8. Duplicates	11	1.75
9. Multiple reasons	46	7.30
Total	630	100

Outcome Measures and Effect Size Extraction

Outcome measures. We chose not to develop rigid operational definitions of the outcome measures, but instead used general descriptions. Achievement outcomes were objective measures—standardized tests, researcher-made or teacher-made tests, or a combination of these—that assessed the extent to which students had achieved the instructional (i.e., learning) objectives of a course. While most measured the acquisition of content knowledge, tests of comprehension and application of knowledge were also included.

Attitude measures and inventories were more subjective reactions, opinions, or expressions of satisfaction, or evaluation of the course as a whole, the instructor, the course

Bernard et al.

content, or the technology used. Some attitude measures could not be classified in these terms and were labeled “other attitudes.”

Retention outcomes were measures of the number or percentage of students who remained in a course out of the total who had enrolled. When these numbers or percentages were expressed in terms of dropout, they were converted to reflect retention.

Effect size extraction. Effect sizes were extracted from numerical or statistical data contained in the study. The basic index for the effect size calculation (d) was the mean of the experimental group (DE) minus the mean of the classroom group divided by the pooled standard deviation (see *Equation 1*).

$$d = \frac{Y_E - Y_C}{S_{Pooled}} \quad (1)$$

Cohen's d was converted to Hedges' g (i.e., unbiased estimate) using Equation 2 (Hedges & Olkin, p. 81).

$$g \cong \left(1 - \frac{3}{4N - 9}\right) d \quad (2)$$

Effect sizes from data in forms such as t -tests, F -tests, p -levels and frequencies were computed via conversion formulas provided by Glass, McGaw, and Smith (1981) and Hedges, Shymansky, and Woodworth (1989). These effect sizes were referred to in coding as “estimated effect sizes.” The following rules governed the calculation of effect sizes:

- When multiple achievement data were reported (e.g., assignments, midterm and final exams, GPAs, grade distributions), the final exam scores were used in calculating the effect size.
- When there was more than one control group and they did not differ considerably, the weighted average of the two conditions was used.
- If only one of the control groups could be considered “purely” control (i.e., classical face-to-face instructional mode), while others involved some elements of DE treatment (e.g., originating studio site), the former was used as the control group.
- In studies where there were two DE conditions and one control condition, the weighted average of the two DE conditions was used.
- In studies where instruction was simultaneously delivered to an originating site and remote sites (e.g. two-way videoconferencing), the originating site was considered to be the control condition and the remote site(s) the DE condition.
- For attitude inventories, we used the average of all items falling under one type of outcome (e.g., attitude towards subject matter) so that only one effect size was generated from each study for each outcome.
- For studies reporting only a significance level, effect sizes were estimated (e.g., $t = 1.96$ for $\alpha = .05$).
- When the direction of the effect was not available, we used an estimated effect size of zero.

Bernard et al.

- When the direction was reported, a “midpoint” approach was taken to estimate a representative t -value (i.e., midpoint between 0 and the critical t -value for the sample size to be significant; Sedlmeier & Gigerenzer, 1989).

The unit of analysis was the independent study finding; multiple outcomes were sometimes extracted from the same study. For within outcome types (e.g., achievement), multiple outcomes were extracted for different courses; when there were several measures for the same course, the more stable outcome (e.g., posttest instead of quizzes) was extracted.

Outcomes and effect sizes from each study were extracted by two researchers, working independently, and then compared for reliability. The inter-coder agreement rate was 91% for the number of effect sizes extracted within studies and 96% for effect size calculation. In total, 688 independent effect sizes (i.e., 321 achievement outcomes, 262 attitude outcomes, and 105 retention outcomes) were extracted.

Study Features Coding

Initial coding. A comprehensive codebook was initially developed based on several earlier narrative reviews (e.g., Phipps & Merisotis, 1999), meta-analyses (e.g., Cavanaugh, 1999), conceptual papers (e.g., Smith & Dillon, 1999), critiques (e.g., Saba, 2000), and a review of 10 sample studies. The codebook was revised as a result of sample coding and a better understanding of the literature and the issues drawn from it. The final codebook included the following categories of study features: outcome features (e.g., outcome measure source), methodology features (e.g., instructor equivalence), course design (e.g., systematic instructional design procedures used), media and delivery (e.g., use of two-way videoconferencing), demographics (e.g., subject matter), and pedagogy (e.g., problem-based learning). Of particular interest in the analysis were the outcomes related to methodology, pedagogy and media characteristics. Some study features were modified and others were dropped (e.g., type of student learning) if there were insufficient data in the primary literature for inclusion in the meta-analysis. The variables and study features that were used in the final coding are contained in the Appendix. In addition to these codes, elaborate operational descriptions were developed for each item and used to guide coders.

Operational definitions of coding options. In order to operationalize the coding scheme and to make coding more concrete, we developed definitions of “more than,” “equal to,” and “less than.” “More than” was defined as 66% or more, “equal to” as 34% to 65%, and “less than” as 33% or less. This approach to coding sets up a comparison between a DE outcome and a control outcome within each coded item which allowed us to quantify some aspects of study features (i.e., methodology, pedagogy, and media) that have heretofore been ignored or dealt with qualitatively. Thus, we hoped the meta-analysis would allow us to address the longstanding controversy regarding the effects of media and pedagogy. As well, this form of coding enabled us to estimate, empirically, the state of the DE research literature from a quality perspective. Each study was coded by two coders independently and compared. Their initial coding agreement was 90%. Disagreements between coders were resolved through discussion and further review of the disputed studies. The whole research team adjudicated some difficult cases.

Synchronous and asynchronous DE. Outcomes were split for the purpose of analysis into synchronous and asynchronous DE on the basis of the study feature “SIMUL.” This study feature described whether the classroom and DE conditions met simultaneously with each other, linked by some form of telecommunication technology such as videoconferencing, or were separate, and therefore not directly linked in any way. The term asynchronous, therefore, does not refer as much to “asynchronous communication” among instructors and/or students as it does

to the fact that it was not synchronized with a classroom. As a result of this definition, some DE students did communicate synchronously with instructors/other students, but this was not typically the case. We did not separate conditions where inter-DE synchronous communication occurred from those where it did not. Outcomes where “SIMUL” was missing were considered “unclassified” and not subjected to thorough analysis (i.e., only their average effect size was calculated).

Recoding methodological study features. Thirteen coded study features relating to the methodological quality of the outcomes were recoded according to the scheme shown in Table 2. Equality between treatment and control was given a weighting of +2 and inequality was recoded as -2 to reflect this extreme discrepancy. The two indeterminate conditions (i.e., one group known and the other not known) were recoded to zero. We had three choices for dealing with the substantial amount of missing information recorded on the coding sheets: 1) use only available information and treat missing data as missing (this would have precluded multiple regression modeling of study features, since each case had at least one study feature missing); 2) recode missing data using a mean substitution procedure under the assumption that missing data were “typical” of the average for each study feature; or 3) code missing data as zero under the assumption that it also represents indetermination. We chose the last of these three options.

Table 2

Methodological Study Feature Codes and Recodes Assigned to Them¹

Study Feature Codes and their Meaning	Recode
1. DE more than control group	-2
2. DE reported/control group not reported	0
3. DE equal to control group	+2
4. Control reported/DE not reported	0
5. DE less than control group	-2
999. Neither DE nor Control Group used explicitly or missing	0

¹Some study had three values (1, 2 and 999) and were coded 2, 1 and 0.

The coded study features were: 1) type of publication; 2) type of measure; 3) effect size (i.e., calculated or estimated); 4) treatment duration; 5) treatment time proximity; 6) instructor equivalence; 7) selection bias; 8) time-on-task equivalence; 9) material equivalence; 10) learner ability equivalence; 11) mortality; 12) class size equivalence; and 13) gender equivalence.

Recoding pedagogical and media study features. To allow us to explore the variability among DE outcomes using multiple regression, we recoded the pedagogical and media-related study features. Using a procedure that is similar to that used to produce the methodological study features, pedagogical and media-related study features were recoded to reflect a contrast between features favoring DE conditions and features favoring classroom conditions. We faced the same problem of missing data with pedagogical and media study features as we did with methodological features. Again we chose to code missing values to zero. Our view was that this was the most conservative approach, since that gave missing values equal weight across all of the study features (i.e., mean substitution would have given unequal weight). An additional reason for favoring this approach was that the bulk of the missing data resided on the classroom side of the scale. This is because, in general, DE conditions were described far more completely than their classroom counterparts. This was especially true for media study features because media represent a definitional criterion of DE, whereas they are not always present in classrooms. So, in effect, many of the relationships expressed in the multiple regression analyses that follow were based on comparisons between a positive value (i.e., either 1 or 2) and 0. Thus, the pedagogical and media study features were recoded using the weighting system shown in Table 3.

Table 3

Pedagogical and Media Study Feature Codes and the Recodes Assigned to Them

Study Feature Codes and their Meaning	Recodes
1. DE more than control group	+2
2. DE reported/control group not reported	+1
3. DE equal to control group	0
4. Control reported/DE not reported	-1
5. DE less than control group	-2
999. Neither DE nor Control Group used explicitly or missing	0

The nine pedagogical coded study features were as follows: 1) systematic instructional design procedures used; 2) advance course information given to students; 3) opportunity for face-to-face (F2F) contact with the teacher; 4) opportunity for F2F contact among student peers; 5) opportunity for mediated communication (e., e-mail, CMC) with the teacher; 6) opportunity for mediated communication among students; 7) student/teacher contact encouraged through activities or course design; 8) student/student contact encouraged through activities or course design; and 9) use of problem-based learning.

The media-related items were as follows: 1) use of two-way audio conferencing; 2) use of two-way videoconferencing; 3) use of computer-mediated communication (CMC); 4) use of e-mail; 5) use of one-way TV or video or audiotape; 6) use of the Web; 7) use of the telephone; and 8) use of computer-based instruction (CBI).

Data Analysis

Aggregating effect sizes. The weighted effect sizes were aggregated to form an overall weighted mean estimate of the treatment effect (i.e., g^+). Thus, more weight was given to findings that were based on larger sample sizes. The significance of the mean effect size was judged by its 95% confidence interval and a z-test. A significantly positive (+) mean effect size indicates that the results favor DE conditions; a significantly negative (−) mean effect size indicates that the results favor traditional classroom-based instruction.

For one study with retention outcomes (Hittleman, 2001) that had extremely large sample sizes (e.g., 1,000,000+), the control sample sizes were reduced to 3,000 with the experimental group's N reduced proportionally. The treatment k was then proportionally weighted. This procedure was used to avoid overweighting by one study. Outlier analyses were performed using the homogeneity statistics reduction method of Hedges and Olkin (1985).

Testing the homogeneity assumption. In addition, Hedges and Olkin's (1985) homogeneity procedures were employed in analyzing the effect sizes for each outcome. This statistic, Q_W , is an extremely sensitive test of the homogeneity assumption that is evaluated using the sampling distribution of χ^2 .

To determine whether the findings for each mean outcome shared a common effect size, the set of effect sizes was tested for homogeneity by the homogeneity statistic (Q_T). When all findings share the same population effect size, Q_T has an approximate χ^2 distribution with $k - 1$ degrees of freedom, where k is the number of effect sizes. If the obtained Q_T value is larger than the critical value, the findings are determined to be significantly heterogeneous, meaning that there is more variability in the effect sizes than chance fluctuation would allow. Study feature analyses were then performed to identify potential moderating factors.

In the study feature analyses, each coded study feature with sufficient variability was tested through two homogeneity statistics: between-class homogeneity (Q_B) and within-class homogeneity (Q_W). Q_B tests for homogeneity of effect sizes across classes. It has an approximate χ^2 distribution with $p - 1$ degrees of freedom, where p is the number of classes. If Q_B is greater than the critical value, it indicates a significant difference among the classes of effect sizes. Q_W indicates whether the effect sizes within each class are homogeneous. It has an approximate χ^2 distribution with $m - 1$ degrees of freedom, where m is the number of effect sizes in each class. If Q_W is greater than the critical value, it indicates that the effect sizes within the class are heterogeneous. Data-analyses were conducted using a meta-analysis software, *Comprehensive Meta-Analysis*TM (Biostat, 2000) and *SPSS*TM (version 11 for the Macintosh OS X, version 10.2.8).

Multiple regression modeling of study features. Weighted multiple regression in SPSS was used to explore the variability in effect sizes and to model the relationships that exist among methodology, pedagogy, and media study features. Each effect size was weighted by the inverse of its sampling variance. Equation 3 was used in calculating the variance, and Equation 4 the weighting factor (Hedges & Olkin, 1985, p. 174).

$$\sigma_d^2 = \frac{n_E + n_C}{n_E n_C} + \frac{d^2}{2(n_E + n_C)} \quad (3)$$

$$W_i = \frac{1}{\sigma_d^2} \quad (4)$$

Each set of study features, methodological, pedagogical and media, was entered into weighted multiple regression separately in blocks using g as the dependent variable and W_i as the weight. Methodology, pedagogy, and media were entered in different orders to assess the relative contribution (R^2 change) of each. Individual methodological, pedagogical, and media study features were then assessed to determine their individual contributions to overall variability. To test the significance of individual study features, the individual β for each predictor was used, and the standard errors were corrected according to Equation 5 (Hedges & Olkin, 1985, p. 174).

$$SE_{Adjusted} = \frac{SE}{\sqrt{MS_E}} \quad (5)$$

The 95% confidence interval was corrected using Equation 6 (Hedges & Olkin, 1985, p. 171).

$$C.I._{Adjusted} = \beta \pm 1.96\sqrt{\sigma_\beta} \quad \text{where, } \sigma_\beta = (SE_{Adjusted})^2 \quad (6)$$

The test statistic z was created for testing the null hypothesis that $\beta = 0$ using Equation 7, and α was evaluated using $t = 1.96$ (Hedges & Olkin, 1985, p. 172).

$$z_{\beta} = \frac{\beta}{\sqrt{\sigma_{\beta}}} \quad (7)$$

Results

In total, 232 studies yielding 688 independent effect sizes (i.e., outcomes) were analyzed. This was based on totals of 57,019 students ($k = 321$) with achievement outcomes, 35,365 students ($k = 262$) with attitude outcomes, and ($N = 57,916,029$) students ($k = 105$) with retention outcomes. The N reported here for retention was reduced to 3,744,869 to avoid overestimation based on a California study of retention over a number of years. The procedure used in reducing these numbers is described in the section on retention outcomes.

Missing Information

One of the most difficult problems we encountered in this analysis was the amount of missing information in the research literature. This, of course, was not a problem for the calculation of effect sizes because the availability of appropriate statistical information was a condition of inclusion. However, it was particularly acute in the coding of study features. Table 4 shows a breakdown of missing study feature data over the three outcome measures: achievement, attitudes, and retention. Overall, nearly 60% of the potentially codable study features were found to be missing. It is because of this difficulty that we recommend caution in interpreting the results based on study features, including methodological quality. Had the research reports been more complete, we would have been able to offer substantially better-quality advice as to what works and what doesn't work in DE.

Table 4
 Number and Percentage of Missing Values in Three Measures

Measure	Total Cells	# Missing	% Missing
Achievement	13,650	7,726	56.61
Retention	4,410	2,664	60.41
Attitude	11,088	5,855	52.80
Total	29,148	16,246	55.73

Achievement Outcomes

Total achievement outcomes. The total number of achievement outcomes was reduced by three outliers, two that exceeded ± 3.0 standard deviations from the mean weighted effect size and one whose Q_w was extreme (i.e., > 500). This left 318 achievement outcomes ($N = 54,775$) to be analyzed. Table 5 shows the frequency and percentage of achievement outcomes classified by their date of publication and Table 6 shows their source of publication.

Table 5
 Date of Publication for Achievement Outcomes

Categories of Publication Date	Frequency	Percentage
1985 – 1989	27	8.49
1990 – 1994	91	28.61
1995 – 1999	108	33.96
2000 – 2002	92	28.93

Table 6
 Categories of Publication for Achievement Outcomes

Categories of Publications	Frequency	Relative Percentage	g^+
Journal articles	135	42.45	-0.009
Dissertations	64	20.13	0.022
Technical Reports	119	37.42	0.036*

$p < .05$.

Based on the data in Table 5, it is clear that the impetus to conduct research of this kind is not diminishing with time, in spite of calls from prominent voices in the field (e.g., Clark, 1983, 1994) that it should. The Pearson Product Moment correlation between year of publication and g is -0.035 ($df = 316$, $p < .05$), indicating that there is no systematic relationship between these two variables. In addition, examination of g^+ in Table 6 indicates that there is modest bias over the three classes of publication sources upon which these data are based. The g^+ for technical reports, while not substantially greater than for dissertations, was significant.

Table 7 shows the weighted mean effect size for 318 outcomes. It is essentially zero, but the test of homogeneity indicates that wide variability surrounds it. This means that the actual average effect size in the population could range substantially on either side of this value.

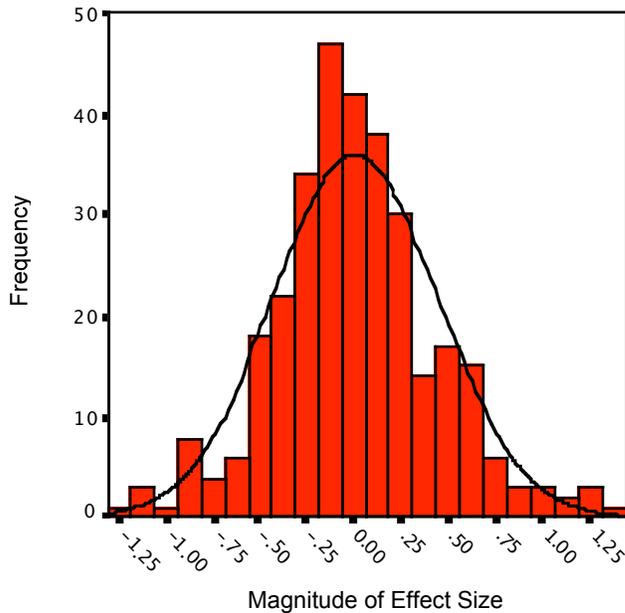
Table 7
 Weighted Mean Effect Size for Combined Achievement Outcomes

Outcomes	Effect Size		95% Confidence Interval		Homogeneity of ES	
	g^+	SE	Lower	Upper	Q -value	df
Combined outcomes ($k = 318$) $N = 54,775$	0.0128	0.0100	-0.0068	0.0325	1191.32*	317

* $p < .05$.

The overall distribution of 318 achievement outcomes is shown in Figure 1. It is a symmetrical distribution with a near zero mean, as indicated, a standard deviation of ± 0.439 , skewness of 0.203, and kurtosis of 0.752; the distribution is nearly normal. It is clear from the range of effect sizes, from -1.31 to $+1.41$, that some applications of DE are far better than classroom instruction and that some are far worse.

Figure 1. Distribution of 318 achievement effect sizes.



Synchronous and asynchronous DE. The split between synchronous and asynchronous DE resulted in 92 synchronous outcomes ($N = 8,677$), 174 asynchronous outcomes ($N = 36,531$) and 52 unclassified outcomes ($N = 9,567$). The mean effect sizes (g^+), standard errors, confidence intervals, and homogeneity statistics for these three categories are shown in Table 8. The difference in g^+ resulting from this split, with synchronous DE significantly negative and asynchronous significantly positive, is dramatic, but both groups remain heterogeneous. Further exploration of the variability in g is required.

Table 8

Weighted Mean Effect sizes for Achievement Outcomes (Synchronous, Asynchronous, and Unclassified)

Categories of DE	Effect Size		95% Confidence Interval		Homogeneity of ES	
	g^+	SE	Lower	Upper	Q -value	df
Synchronous DE ($k = 92$) $N = 8,677$	-0.1022*	0.0236	-0.1485	-0.0559	182.11*	91
Asynchronous DE ($k = 174$) $N = 36,531$	0.0527*	0.0121	0.0289	0.0764	779.38*	173
Unclassified DE ($k = 52$) $N = 9,567$	-0.0359	0.0273	-0.0895	0.0177	191.93*	51

* $p < .001$

Weighted multiple regression. In beginning to explore the variability in g , we conducted weighted multiple regression (WMR) with the three blocks of predictors. We were particularly interested in the variance accounted for by each of the blocks—methodology, pedagogy, and media—entered in different orders to determine their relative contribution to achievement. Clark and others have argued that poor methodological quality tends to confound the effect attributable to features of pedagogy and media, and that pedagogy and media themselves are confounded in

studies of this type. In this analysis, we have attempted to untangle these confounds and to suggest where future researchers and designers of DE applications should expend their energy. WMR was used to assess the relative contributions of these three blocks of predictors. The weighting factor, as described in the Method section, is the inverse of the variance and the dependent variable in all cases was g (Hedges & Olkin, 1985).

We begin with an overall analysis, followed by a more detailed, albeit more speculative, description of the particular study features that account for the more general findings. We entered the three blocks of predictors¹ (e.g., 13 methodological study features) into WMR in different orders: 1) methodology, then pedagogy, then media; 2) methodology, then media, then pedagogy; 3) pedagogy, then media, then methodology; and 4) media, then pedagogy, then methodology. We did not enter methodology on the second step because this combination seemed to explain little of interest. Table 9 shows the partitioning of variance between (Q_B) and within (Q_W) on the third step of regression for both synchronous and asynchronous DE outcomes. These statistics are applicable for all of the orders. Q_B is significant for both DE patterns, and synchronous DE outcomes are homogeneous (i.e., Q_W is not significant), while asynchronous DE outcomes are not (i.e., Q_W is significant). The critical value of χ^2 at each appropriate df was used to test the significance of each effect.

Table 9

Tests of Between and Within-group Variation for Synchronous and Asynchronous Achievement Outcomes

Source	Synchronous DE		Asynchronous DE	
	<i>SS</i>	<i>df</i>	<i>SS</i>	<i>df</i>
Q_B	111.32*	26	222.41*	30
Q_W	66.96	65	548.84*	143
Total	178.29	91	771.25	173

* $p < .05$.

Table 10 provides a comparison of the R^2 changes for each of the blocks of predictors. This table reveals some interesting insights into the nature of these predictors, relative to one another. First, with one exception each (i.e., third step in both cases), methodology and pedagogy are always significant, no matter which position they are in or whether the outcomes are associated with synchronous or asynchronous DE. Second, only when media is entered on the first step is it significant. Overall, this says that methodology and pedagogy are more important than media as predictors of achievement. Third, in line with much of the commentary on the research literature of DE and other media comparison literatures, research methodology accounts for a substantial proportion of variation in effect size, more for synchronous than for asynchronous DE. One of the difficulties with previous meta-analyses of these literatures is that, at best, methodologically unsound studies were removed *a priori*, often by fuzzy criteria such as “more than one methodological flaw.” By including studies that range in methodological quality, and coding for it, we have overcome this difficulty to an extent.

Table 10
 Comparison of R^2 Change for Blocks of Study Features for Achievement Outcomes

Predictors	1st Step	2nd Step after Methodology	2nd Step after Pedagogy/Media	3rd Step
Synchronous DE				
Methodology	0.490*			0.250*
Pedagogy	0.360*	0.101*	0.130**	0.077
Media	0.245*	0.058	0.015	0.048
Asynchronous DE				
Methodology	0.117*			0.054
Pedagogy	0.156*	0.107*	0.124*	0.120*
Media	0.111*	0.051	0.078	0.065

* $p < .05$. Note: Not all significance tests are based on the same degrees of freedom. - - -

Finally, we calculated the predicted g^+ that resulted after each step of WMR, with methodology entered first, pedagogy entered second, and media entered last. The statistics in Table 11 are the equivalents of the predicted means in *ANCOVA* (i.e., $Y_{\text{Predicted}}$) and are the expected values if each of these blocks could have been controlled for in the original experiments and can be compared to the unadjusted g^+ in row one of the table. Interestingly, g^+ for synchronous outcomes increases and g^+ for asynchronous outcomes decreases.

Table 11
 Actual and Predicted g^+ at for Methodology, Pedagogy, and Media Achievement Outcomes

Blocks of Predictors	Synchronous DE	Asynchronous DE
	g^+	g^+
Unadjusted	-0.1022	0.0527
After Methodology	-0.0599*	0.0328*
After Pedagogy	-0.0145*	0.0238*
After Media	-0.0555*	0.0425*

* Predicted g^+ .

Study feature analysis. We now proceed to a more detailed analysis of study features that were outcomes of the previous analysis. A logical way of approaching this is to present the results with methodology on the first step and then both pedagogy and media on separate WMR runs on the second step. This allows for an assessment of pedagogy and media, independently, after variation in methodology has been removed. Therefore, as shown in the second column in Table 10, these results derive from pedagogy and media.

Table 12 shows the individual study feature results for synchronous DE outcomes and Table 13 shows the same results for asynchronous DE outcomes. Shown are the original betas and standard errors along with the adjusted standard errors (see Equation 5), the adjusted confidence intervals (see Equation 6), and the adjusted z -tests (see Equation 7), evaluated with $t_{\text{Critical}} = 1.96$ at a probability of .05.

Table 12
Significant Individual Study Features for Synchronous Achievement Outcomes

Predictors	β	SE	$SE_{adj.}$	Confidence Intervals		z-test
				Lower _{adj.}	Upper _{adj.}	
Pedagogy Study Features (Step 2 after methodology)						
Face-to-face with instructor	-0.239	0.074	0.07249 ²	-0.3810	-0.0969	-3.29*
Face-to-face with Students	0.183	0.073	0.07151	0.0428	0.3231	2.55*
Media Study Features (Step 2 after methodology)						
TV-Video	0.186	0.092	0.08751 ³	0.0144	0.3575	2.12*
Telephone	-0.157	0.065	0.06183	-0.2781	-0.0358	-2.54*

* $p < .05$, $t_{critical} = 1.96$, ¹MS = 1.166, ²MS = 1.042, ³MS = 1.105

Table 13
Significant Individual Study Features for Asynchronous Achievement Outcomes

Predictors	β	SE	$SE_{adj.}$	Confidence Intervals		z-test
				Lower _{adj.}	Upper _{adj.}	
Pedagogy Study Features (Step 2 after methodology)						
Advance Course Information	0.120	0.047	0.024	0.074	0.166	5.08*
Mediated Comm. with Instructor	0.128	0.057	0.029	0.072	0.184	4.47*
Problem-based Learning	0.280	0.145	0.073	0.137	0.423	3.84*
Media Study Feature (Step 2 after methodology)						
TV-Video	0.124	0.069	0.034 ³	0.058	0.190	3.69*

* $p < .05$. $t_{critical} = 1.96$, ¹MS = 4.256, ²MS = 3.966, ³MS = 4.222.

Demographic study features. We also coded a set of study features relating to the demographics of students, instructors, subject matter and reasons for offering DE. Table 14 contains the three study features which yielded enough outcomes to warrant analysis. DE achievement effects were large when: a) the efficient delivery or cost was a reason for offering DE courses ($g+ = 0.1639$); b) for students in grades k-12 ($g+ = 0.2016$); and c) for military and business subject matters ($g+ = .1777$). Interestingly, there was no difference between undergraduate post-secondary education applications of DE and classroom instruction. Graduate school applications yielded modest but significant results in favor of DE ($g+ = 0.0809$). As well, the academic subject areas of math, science and engineering appear to be best suited to the classroom ($g+ = -0.1026$), while subjects related to computing and the military/business ($g+ > 0.17$) seem to work well in distance education settings.

Table 14
 Effect Sizes for Demographic Study Features ($k \geq 10$)

Study Features	g^+	t -value
Reasons for offering DE courses		
Access to expertise ($k = 48$)	-0.0821	-2.93**
Efficient delivery or cost ($k = 22$)	0.1639	3.55**
Multiple purposes ($k = 22$)	0.1557	2.84**
Type of students		
K-12 ($k = 24$)	0.2016	4.26**
Undergraduate ($k = 219$)	-0.0048	-0.38
Graduate ($k = 36$)	0.0809	2.18*
Military ($k = 11$)	0.4452	6.80**
Subject matter		
Math, science and engineering ($k = 67$)	-0.1026	-3.94**
Computer science/computer applications ($k = 13$)	0.1706	3.01**
Military/business ($k = 50$)	0.1777	5.72**

* $p \leq .05$. ** $p < .01$.

Attitude Outcomes

Synchronous and asynchronous outcomes. We found various forms of attitude measures in the literature that could be classified into four categories: attitude towards technology; attitude towards subject matter; attitude towards instructor; and attitude towards course. We also have a fairly large set of measures ($k = 90$) that could not be classified into a single set and we therefore labeled them as “Other Attitude Measures.” We chose not to include “Other Attitudes” in the same analysis where the type of measure was known. Therefore, the total number of attitude outcomes was reduced from 262 to 172. This number was further reduced when missing data prevented us from categorizing outcomes as either synchronous or asynchronous. Before analysis, one extremely high outlier was removed. This left 154 outcomes to be analyzed.

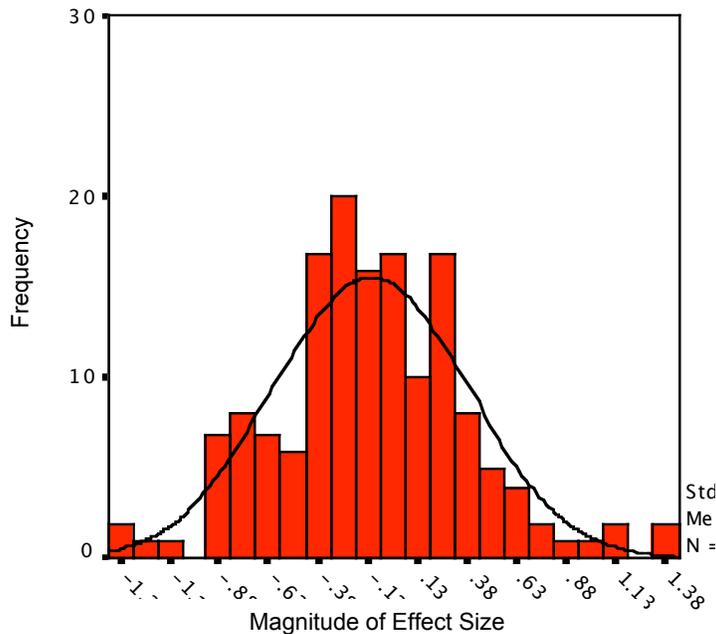
We split the sample into synchronous and asynchronous DE, in the same manner as for achievement, and found essentially the same overall dichotomy. Table 15 shows these results along with the results of 154 combined attitudes (i.e., before classification into synchronous and asynchronous). While all of the weighted mean effect sizes are negative, notice the contrast between synchronous and asynchronous outcomes. The average effect size for synchronous outcomes is significant, while the average effect size for asynchronous is not. Furthermore, there is high variability among effect sizes, even after the split. Figure 2 provides a graphic depiction of overall variability in attitude outcomes for 154 outcomes and shows that they range from +2.41 to -2.38. There are circumstances where DE student reactions are extremely positive and others where reactions are quite negative, relative to classroom instruction.

Table 15
 Weighted Mean Effect Sizes for Combined, Synchronous, and Asynchronous Attitude Outcomes

Categories of DE	Effect Size		95% Confidence Interval		Homogeneity of <i>ES</i>	
	<i>g</i> ⁺	<i>SE</i>	Lower	Upper	<i>Q</i> -value	<i>df</i>
Combined (not including "other attitudes" (<i>k</i> = 154) <i>N</i> = 21,047)	-0.0812*	0.0146	-0.1098	-0.0526	793.65*	153
Synchronous DE (<i>k</i> = 83) <i>N</i> = 9,483	-0.1846*	0.0222	-0.2282	-0.1410	410.02*	82
Asynchronous DE (<i>k</i> = 71) <i>N</i> = 11,624	-0.0034	0.0193	-0.0412	0.0344	345.64*	70

**p* < .001.

Figure 2. Distribution of 154 attitude effect sizes.



Weighted multiple regression. Given the wide variability in attitude outcomes, a WMR analysis was conducted in a manner similar to the one done with the achievement data. The within-group and between-groups tests of significance are shown in Table 16. *Q_W* is significant for synchronous and asynchronous DE outcomes, indicating heterogeneity for these groups.

Table 16
Tests of Between- and Within-group Variation for Synchronous and Asynchronous Attitude Outcomes

Source	Synchronous DE		Asynchronous DE	
	<i>SS</i>	<i>df</i>	<i>SS</i>	<i>df</i>
Q_B	260.68*	22	154.98*	26
Q_W	135.35*	60	136.99*	44
Total	396.025	82	771.25	70

* $p < .05$.

We examined the R^2 change for three blocks of predictors: methodology, pedagogy, and media, for attitudes in different orders, in the same way we did for achievement outcomes. Table 17 is a comparison of R^2 change for blocks of study features entered in different orders in WMR. The results do not as clearly favor methodology, pedagogy, and the diminished role of media as they did for achievement. In fact, these results indicate a more complex relationship among the three blocks of predictors. For one thing, there are more differences here between synchronous and asynchronous DE for the three blocks of predictors. As with achievement, methodology still accounts for more variation in synchronous DE than in asynchronous DE. While pedagogy is somewhat suppressed for synchronous DE, it emerges as important for asynchronous DE. On the other hand, media appears to be more important in synchronous DE than in asynchronous DE.

Table 17
Comparison of R^2 Change for Blocks of Study Features for Attitude Outcomes

Predictors	1st Step	2nd Step after Methodology	2nd Step after Pedagogy/Media	3rd Step
Synchronous DE				
Methodology	0.471**			0.421**
Pedagogy	0.128	0.138**	0.101	0.120**
Media	0.136**	0.067*	0.109**	0.049
Asynchronous DE				
Methodology	0.218**			0.157
Pedagogy	0.253**	0.215**	0.133	0.076
Media	0.241**	0.236**	0.121	0.097

* $p = .057$. ** $p < .05$. Note. Not all significance tests are based on the same degrees of freedom.

Table 18 shows the predicted $g+$ results before WMR and after methodology, pedagogy, and media were entered. Synchronous DE becomes more negative (i.e., favoring the classroom condition) and asynchronous DE moves from an initial negative sign to a positive sign after methodology is entered, and thereafter.

Table 18
Predicted $g+$ for Methodology, Pedagogy, and Media for Attitude Outcomes

Blocks of Predictors	Synchronous DE	Asynchronous DE
	$g+$	$g+$
Unadjusted	-0.1846	-0.0034
After Methodology	-0.1707*	0.0243*
After Pedagogy	-0.1702*	0.0212*
After Media	-0.2031*	0.0237*

* Predicted $g+$

Study feature analysis. Individual study features were assessed after WMR in a manner similar to that for achievement outcomes; the significant results are shown in Table 19 for synchronous DE outcomes and in Table 20 for asynchronous DE outcomes. Again, regression information from methodology on the first step, pedagogy on the second step (i.e., after methodology), and media on the second step (i.e., after methodology) is presented. The adjustments to standard errors, confidence intervals, and z-tests were performed according to the equations in the Method section.

Table 19
 Individual Study Features for Synchronous Attitude Outcomes

Predictors	β	SE	SE _{adj.}	Confidence Intervals		z-test
				Lower _{adj.}	Upper _{adj.}	
Pedagogy Study Features (Step 2 after methodology)						
Advance Info.	-0.281	0.138	0.089 ²	-0.455	-0.107	-3.17*
Systematic ID	1.162	0.385	0.248	0.677	1.647	4.69*
Mediated comm. with instructor	0.061	0.255	0.164	0.279	0.923	3.66*
Instructor/student contact encouraged	0.314	0.135	0.087	0.144	0.484	3.62*
Media Study Features (Step 2 after methodology)						
TV-Video	-0.351	0.126	0.076 ³	-0.500	-0.202	-4.60*
Use of telephone	0.262	0.103	0.062	0.140	0.384	4.203*

* $p < .05$. $t_{critical} = 1.96$, ¹MS = 2.949, ²MS = 2.416, ³MS = 2.73.

Table 20
 Individual Study Features for Asynchronous Attitude Outcomes

Predictors	β	SE	SE _{adj.}	Confidence Intervals		z-test
				Lower _{adj.}	Upper _{adj.}	
Pedagogy Study Feature (Step 2 after methodology)						
Problem-based learning	0.403	0.180	0.099 ²	0.209	0.597	4.07*
Media Study Features (Step 2 after methodology)						
CMC	0.272	0.108	0.099 ³	0.151	0.373	4.41*
CBI	0.392	0.150	0.086	0.224	0.560	4.57*
Web	-0.168	0.191	0.520	-0.270	-0.066	-3.23*

* $p < .05$. $t_{critical} = 1.96$, ¹MS = 3.934, ²MS = 3.307, ³MS = 3.063.

Retention Outcomes

Retention is defined here as the opposite of dropout or attrition. We found several studies of statewide data (i.e., California) that compared DE to classroom conditions, where the sample size was in the millions. To correct for the extreme effects of these huge ($N = 57,916,029$), but anomalous studies, we truncated the sample sizes of the classroom condition to 3,000 and proportionately reduced the DE condition to create a better balance with other studies ($N = 3,735,050$). Otherwise, these effect sizes would have dominated the average effects, unduly skewing it in favor of the large samples. Figure 3 shows the distribution of effect sizes for the retention measure. The distribution is clearly bimodal, with the primary mode at zero. Again, there is wide variability.

Figure 3. Distribution of 70 retention effect sizes.

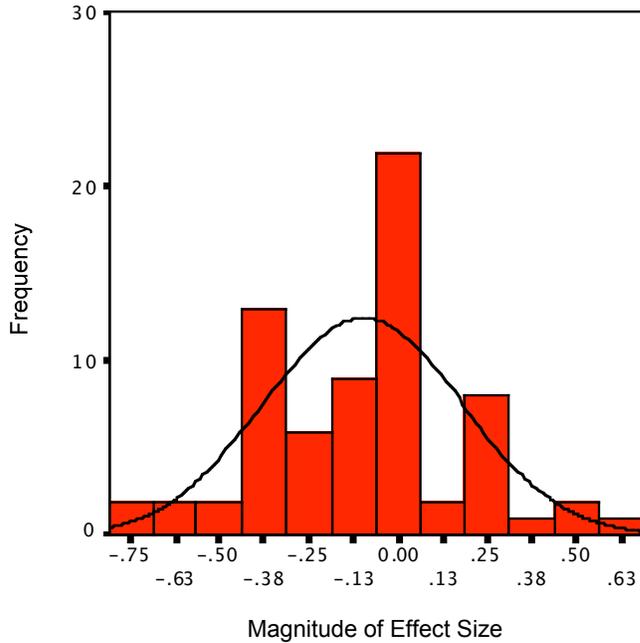


Table 21 shows the results of this analysis and the results of the split between synchronous and asynchronous DE conditions. None of the large-sample studies had been coded as either synchronous or asynchronous and so, while the number of effects is fairly representative of the total, the number of students is not. In spite of this, the results of the synchronous/asynchronous split seem to reflect the average for all studies. Caution in the interpretation of the mean effect size for synchronous DE should be exercised because of the low number of outcomes associated with it.

Table 21
 Mean Effect Sizes for Synchronous and Asynchronous Retention Outcomes

Outcome Type	Effect Size		95% Confidence Interval		Homogeneity of <i>ES</i>	
	<i>g</i> +	<i>SE</i>	Lower	Upper	<i>Q</i> -value	<i>df</i>
Overall Retention (<i>k</i> = 103) <i>N</i> = 3,735,050	-0.0573*	0.0065	-0.0700	-0.0445	3150.96*	102
Synchronous DE (<i>k</i> = 17) <i>N</i> = 3,604	0.0051	0.0341	-0.0617	0.0718	17.17	16
Asynchronous DE (<i>k</i> = 53) <i>N</i> = 10,435	-0.0933*	0.0211	-0.1347	-0.0519	70.52*	52

**p* < .05.

Since the traditionally high dropout rate in DE has been attributed to factors such as isolation, poor student-teacher communication, etc., we wondered if this situation had changed over the years of this study, with the increasing availability newer forms of electronic communication. To explore this, we calculated the correlation between dropout (i.e., *g*) and “year of publication” over the 17 years of the study. The Pearson Product-Moment Correlation is 0.015 (*df* = 68, *p* > .05), suggesting that there is no systematic increase or decrease in the differential retention rate over time. This situation was somewhat different for synchronous (*r* = -0.27, *df* =

Bernard et al.

14, $p > .05$) and asynchronous retention outcomes ($r = 0.011$, $df = 51$, $p > .05$), calculated separately, although neither reached significance. Had the synchronous correlation been significant, it would have indicated a decreasing differential (i.e., the two conditions becoming more the same) over time between classroom and DE applications in terms of retention.

When WMR was performed on synchronous and asynchronous retention outcomes, neither the results of methodology, pedagogy, or media were significant. Therefore, no regression outcomes are presented.

Summary of Results: Achievement

1. There is a very small and significant effect favoring DE conditions ($g+ = 0.0128$) on overall achievement outcomes ($k = 318$). However the variability surrounding this mean is wide and significant.
2. When outcomes were split between synchronous and asynchronous DE achievement outcomes, a small significant negative effect ($g+ = -0.1022$) occurred for synchronous DE and a significantly positive effect occurred for asynchronous DE ($g+ = 0.0527$). Variability remained wide and significantly heterogeneous for each group.
3. WMR revealed that together, methodology, pedagogy, and media accounted for 62.4% of variation in synchronous DE achievement outcomes and 28.8% of variability in asynchronous DE outcomes.
4. When R^2 change was examined for blocks of predictors, entered in different orders, methodology and pedagogy were almost always found to be significant, whereas media was only significant when it was entered on the first step. This was true for both synchronous DE and asynchronous DE outcomes. Individual significant study feature outcomes are summarized in Table 22.

Summary of Results: Attitude

1. There is a small negative but significant effect in favor of classroom instruction ($g+ = -0.0812$) on overall attitude outcomes. Again, the variability around this mean is significantly heterogeneous.
2. There are differences in effect sizes for synchronous DE ($g+ = -0.1846$) and asynchronous DE ($g+ = -0.0034$). Both favor classroom instruction but the average effect size is significant for synchronous DE and it is not for asynchronous DE. Individual significant study feature outcomes are summarized in Table 22.
3. R^2 change analysis of the type described above revealed varying patterns of variance, accounted for by methodology, pedagogy, and media in terms of attitudes. It appears that these three sets of variables are related in a more complex way than they are achievement outcomes.

Summary of Results: Retention

1. There is a very small but significant effect in favor of classroom instruction ($g+ = -0.0573$) on retention outcomes.
2. There is a very small but positive effect for synchronous DE, which is not significant ($g+ = 0.0051$), and a larger negative effect ($g+ = -0.0933$) for asynchronous DE.

Summary of Results: Overall

1. There is extremely wide variability in effect size on all measures, and we were unable to find study features that form homogeneous subsets, including the distinction between synchronous and asynchronous DE (with the one exception of synchronous DE on achievement). This suggests that DE works extremely well sometimes and extremely poorly other times, even when all coded study features are accounted for.
2. Since the variation in effect size accounted for by methodology is fairly substantial (generally speaking, more substantial for synchronous than asynchronous DE), and often more so than for pedagogy and media combined, methodological weakness was considered an important deterrent to forming clear recommendations to practitioners and policymakers.
3. Another measure of the quality of the literature, amount of data available, suggests that the literature is very weak in design features that would improve the interpretability of the results. Over half (55.73%) of the codable study features (including methodological features) were missing.
4. Even though the literature is large, it is difficult to draw firm conclusions on what works and doesn't work in DE, except to say that the distinction between synchronous and asynchronous forms of DE does moderate effect sizes in terms of both achievement and attitudes. Concise statements of outcomes based on study feature analysis (Table 22) are made with caution and remain speculative because of relatively large amount of missing data relating to them.

Table 22 is a summary of the significant study features that resulted from WMR.

Table 22

Summary of Study Features that Significantly Predict Achievement, Attitude, and Retention Outcomes

Synchronous DE	
Favor Classroom Instruction (-)	Favor Distance Education (+)
<i>Achievement</i>	
<ul style="list-style-type: none"> • Face-to-face meetings with the instructor • Use of the telephone to contact instructor 	<ul style="list-style-type: none"> • Face-to-face contact with other students • Use of one-way TV-video
<i>Attitudes</i>	
<ul style="list-style-type: none"> • Opportunity for face-to-face contact with other students • Use of one-way TV-video 	<ul style="list-style-type: none"> • Use of systematic ID • Opportunity for mediated communication with the instructor • Instructor/student contact encouraged • Use of the telephone to contact instructor
<i>Retention</i>	
<ul style="list-style-type: none"> • No significant predictors 	<ul style="list-style-type: none"> • No significant predictors

Asynchronous DE	
Favor Classroom Instruction (-)	Favor Distance Education (+)
<u>Achievement</u> <ul style="list-style-type: none"> • No significant predictors 	<u>Achievement</u> <ul style="list-style-type: none"> • Use of problem-based learning strategies • Opportunity for mediated communication with the instructor • Advance information given to students • Use of one-way TV-video
<u>Attitudes</u> <ul style="list-style-type: none"> • Use of the Web 	<u>Attitudes</u> <ul style="list-style-type: none"> • Use of problem-based learning strategies • Use of computer-mediated communication • Use of computer-based instruction
<u>Retention</u> <ul style="list-style-type: none"> • No significant Predictors 	<u>Retention</u> <ul style="list-style-type: none"> • No significant predictors

Discussion

Overall Findings

The most important outcome of the overall analysis of effect size relates to the wide variability in outcomes for all three primary measures. While the average effect of DE was near zero, there is a tremendous range of effect sizes (*g*) in achievement outcomes—from +1.41 to -1.31. There are instances where the DE group outperformed the traditional instruction group by more than 50%. And there are instances where the opposite occurred—the traditional instructional group outperformed the DE group by 48% or more. This is likewise for overall attitude and retention outcomes.

None of the measures is homogeneous, so interpreting means as if they are true representations of population values is risky (Hedges & Olkin, 1985). It is simply incorrect to say that DE is better than, worse than, or even equal to classroom instruction on the basis of mean effect sizes and heterogeneity. This wide variability means that a substantial number of DE applications provide better achievement results, are viewed more positively and have higher retention rates than their classroom counterparts. On the other hand, a substantial number of DE applications are far worse than classroom instruction on all three measures.

The mistake that a number of previous reviewers have made, from early narrative reviews (e.g., Moore & Thompson, 1990) to more recent reviews (e.g., Russell, 1999), is to declare that DE and classroom instruction are equal without examining the variability surrounding their difference. Wide and unexplained variability precludes any such simplistic conclusion. An assessment of the literature of this sort can only be made through a meta-analysis that provides a comprehensive representation of the literature, the application of rigorously applied inclusion/exclusion criteria, and an analysis of variability around mean effect sizes. On a further note, the overall retention outcomes appear to indicate that the substantial degree of retention

Bernard et al.

differential between classroom and DE conditions, noted in many studies of student persistence, is still present in these studies.

Quality of the DE Literature

In the last few years, a number of commentators (Anglin & Morrison, 2000; Diaz, 2000; Perraton, 2000; Saba, 2000; Phipps & Merisotis, 1999) have decried the quality of the DE research literature. One of the main purposes of this meta-analysis was to estimate the extent of these claims and to examine the depth of the research literature in terms of its completeness. This discussion begins with that assessment, because both the quality of studies and the depth of reporting impinge upon all other aspects of the analysis.

One whole section of the codebook (13 items) deals with methodological aspects of the studies that were reviewed. Our intent was not to exclude studies that had methodological weaknesses, such as lack of random assignment or non-equivalent materials, but to code these features and examine how they affect the conclusions that can be drawn from the studies. However, the quality and quantity of reporting in the literature that we examined affect the accuracy of the methodological assessment, since missing aspects of design, control, measurement, equivalence of conditions, etc. influence the quality of the methodological assessment.

Information available in the literature. Overall, we found the literature severely wanting in terms of depth of reporting. Nearly 60% of the codable study features, including methodological features, were coded as missing. This means that for outcomes that met our inclusion criteria and for which we could calculate an effect size, we were only able to derive a 40% estimate of the study features on the effect sizes. The most persistent problem was the reporting of characteristics of the comparison condition (i.e., classroom instruction). Often, authors went to extraordinary lengths to describe the DE condition, only to say that it was being compared to a “classroom condition.” If we cannot discern what a DE condition is being compared to, it is very difficult to come to any conclusion as to what an effect size characterizing their difference means. This was not just a problem in reports and conference papers that are often not reviewed or reviewed only at a cursory level; it was true of journal articles and dissertations, as well, which are presumably reviewed by panels of peers or committees of academics. This speaks not only to the quality of the peer review process of journals but to the quality and rigor of training that future researchers in our field are receiving. However, an analysis of publication source revealed only a small bias in mean effect size among the types of literature that are represented in these data (i.e., achievement data only).

There are some interesting statistics associated with year of publication that bear noting. In spite of calls from the field to end the form of classroom comparative studies investigated here (e.g., Clark, 1983, 1994), their frequency actually appears to have been increasing since 1985. As indicated in the Results section, there appears to be no systematic relationship between “year of publication” and effect size.

Methodological quality of the literature. Field experiments investigating educational practices are characteristically weak because they are so often conducted in circumstances where the opportunities to control for rival explanations of research hypotheses are minimal. Therefore, they are typically higher in external validity than in internal validity. Cook and Campbell (1979) argue that this trade-off between internal and external validity is justified under certain circumstances. The What Works Clearinghouse (Valentine & Cooper, 2003) uses a four-axis model of research methodology, based on Shadish, Cook, and Campbell (2002), to judge the quality of a research study: internal validity, external validity, measurement validity, and statistical validity. Our 13 coded study features relating to methodology focused more on internal

validity than on the other three types of validity. Ten items rated aspects of internal validity in terms of the equality or inequality of comparison groups; no direct assessment of external validity was made, one feature assessed the quality of the outcome measure used, another assessed the quality of the publication source, and another rated the quality of the statistical information used in calculating effect sizes (i.e., calculated or estimated).

Remembering that many codable aspects of methodological quality were unavailable because of missing information, we tried to characterize the quality of studies in terms of research design and the degree of control for confounding. We chose to enter the 13 methodological study features into weighted multiple regression as a way of: 1) assessing methodology independently and in relation to other blocks of study features, and 2) assessing other study features after variation due to methodology was removed. We found that methodology accounts for a substantial proportion of overall variation in the effect sizes for achievement and attitude measures. This was moderated somewhat when outcomes were split between synchronous and asynchronous DE patterns. Typically, more methodological variation was accounted for in synchronous DE than in asynchronous DE.

Our recoding scheme emphasized the difference between methodological strengths and methodological weaknesses, with missing data considered neutral. In a strong experimental literature, with little missing data, strong measures, and adequate control over confounding, the variance accounted for by methodology would have been minimal. In the most extreme case, zero variability would be attributable to methodology. As previously indicated, this was not the case, suggesting that the dual contributing factors of experimental and methodological inadequacies and missing information weaken this DE research literature. However, this fact does not mitigate entirely against exploring these data in an effort to learn more about the characteristics of DE and the relative contributions of various factors to its success or failure, relative to classroom instruction.

Synchronous and Asynchronous DE

After assessing overall outcomes for the three measures, we split the samples into the two different forms of DE noted in the literature, synchronous DE and asynchronous DE. Synchronous DE is defined by the time- and place-dependent nature of classroom instruction proceeding in synchronization with a DE classroom, located in a remote location, and connected by videoconferencing, audio-conferencing media, or both. Asynchronous DE conditions were run independently of their classroom comparison conditions. While a few asynchronous applications actually used synchronous media among themselves, they were not bound by time and place to the classroom comparison condition. The current use of the word asynchronous often refers to the lag-time in communication that distinguishes, for instance, e-mail from a “chat room”; our definition does not disqualify some synchronous communication between students and instructors and students and other students.

The results of this split yielded substantially different outcomes for the two forms of DE on all three measures. In the case of achievement, synchronous outcomes favored the classroom condition, ranging from +0.97 to -1.14 (this is the only homogeneous subset), while asynchronous outcomes favored the DE condition, ranging from +1.41 to -1.31. While both mean effect sizes for attitudes were negative, the differences were dramatically different for synchronous and asynchronous DE, favoring classroom instruction by nearly 0.20 *SD*. The split for retention outcomes yielded the opposite outcome. Dropout was substantially higher in asynchronous DE, compared with synchronous DE.

It is possible that these three results can be explained in the same terms by examining the conditions under which students learn and develop attitudes in these two patterns, as well as

Bernard et al.

make decisions to persist or drop out. Looked at in one way, synchronous DE is a poorer-quality replication of classroom instruction; there is neither the flexibility of scheduling and place of learning nor the individual attention that exists in many applications of asynchronous DE, and there is the question of the effectiveness of “face-to-face” instruction conducted through a teleconferencing medium. Although we were unable to ascertain much about teaching style from the literature, there may be a tendency for synchronous DE instructors to engage in lecture-based, instructor-oriented strategies that may not translate well to mediated classrooms at a distance (Verduin & Clark, 1991). Even employing effective questioning strategies may be problematic under these circumstances. In fact, there have been calls in the literature of synchronous DE for instructors to adopt more constructivist teaching practices (Beaudoin, 1990; Dillon & Walsh, 1992; Gehlauf, Shatz, & Frye, 1991). According to Bates (1997) asynchronous DE, by contrast, can more effectively provide interpersonal interaction and support two-way communication between instructors and students and among students, thereby producing a better approximation to a learner-centered environment. These two sides of the DE coin may help explain the differential achievement and attitude results.

Work carried out by Chickering and Gamson (1987) offers an interesting framework to address the question of teaching in DE environments. Based on 50 years of higher education research, they produced a list of seven basic principles of good teaching practices in face-to-face courses. Graham, Cagiltay, Craner, Lim, and Duff (2000) used these same seven principles to assess whether these skills transfer to online teaching environments. Their general findings, echoed by the work of Schoenfeld-Tacher and Persichitte (2000) and Spector (2001), indicate that DE teachers typically require different sets of technical and pedagogical competencies to engage in superior teaching practices, although Kanuka, Collett, & Caswell (2003) claim that this transition can be made fairly easily by experienced instructors. Presumably, this applies to both synchronous and asynchronous DE, but because synchronous DE is more like classroom instruction, and is performed in view of a live classroom as well as a mediated one, it is possible that adopting new and more appropriate teaching methods is not as critical and pressing as it is in asynchronous DE.

If achievement is better and attitudes are more positive in asynchronous DE than in synchronous DE, why is its retention rate lower? First of all, based on the literature, it is not surprising that there is greater dropout in DE courses than in traditional classroom-based courses (Kember, 1996). The literature has said this for years. However, this does not fully answer the question about synchronous and asynchronous DE. Part of the answer is that achievement and attitude measurement are independent of retention since they do not include data from students who dropped out before the course ended. A second part of the answer may lie, again, in differences in the conditions that exist in synchronous and asynchronous DE. As previously noted, synchronous DE is more like classroom instruction than is asynchronous DE. Students meet together in a particular place, at a particular time. They are a group, just like the classroom students. The difference is that they are remote from the instructor. Students working in asynchronous DE conditions do not typically meet in groups, although they may have face-to-face and/or synchronous mediated contact with the instructor and other students. Group affiliation and social pressure, then, may partially explain this effect. Other explanations may derive from models of persistence such as Kember's (1996), which stress factors such as entry characteristics, social integration, external attribution, and academic integration.

Only a small percentage of the findings for synchronous DE are based on K-12 learners. We speculate that for younger learners the structure of synchronous DE may be better suited to their academic schedules and their need for spontaneous guidance and feedback. Furthermore, we have concerns about the nature of appropriate comparisons. For example, how does

asynchronous DE compare to home schooling or the provision of specialized content by a non-expert (e.g., in rural and remote communities)?

This question is an even more general concern that goes beyond synchronicity or asynchronicity of DE delivery and addresses the question of access to education and the appropriate nature of the comparison condition. When is it appropriate for DE to be compared to traditional instruction, other alternative delivery methods, or a no instruction control group? In the latter case, this may be the choice with which a substantial number of learners are faced and which represents one purpose of DE—to provide learning opportunities when no others exist. In such circumstances, issues of instructional quality, attitudes, and retention may be secondary to issues of whether assessment and outcome standards—ensuring rigorous learning objectives—are maintained.

Media vs. Pedagogy: Resolving the Debate?

Is technology transparent or is it transformative? Do the most effective forms of DE take unique advantage of communication and multimedia technologies in ways absent from “traditional” classroom instruction? If so, why are these absent from classroom instruction? For example, how much does the distance-learning context provide the requisite incentive for learners to use the technological features apparent in some media-rich DE applications? Alternatively, can effective pedagogy exist independently of the advantages and restrictions of DE? Can, for example, clarity, expressiveness, and instructional feedback be provided regardless of their medium of delivery and independently of the separation of space and time? Finally, how can we begin to explore these issues independently of concerns about methodological quality and completeness?

The nature of the DE research literature, in which research methodology, pedagogy, and media are all present and intertwined, gave us an opportunity to examine their relative contributions to achievement, attitude, and retention outcomes and to further explore the wide variability that still existed after studies were split into synchronous and asynchronous DE. We settled on an approach to weighted multiple regression (WMR) in which blocks of these recoded study features were entered in different orders and assessed the R^2 change that resulted from their various positions in the regression models. With the exception of retention, which did not achieve statistical significance for either type of DE, the overall percentage of variance accounted for by these blocks ranged from 29% to 66% for achievement and attitude. However, only one homogeneous set was found: achievement outcomes for synchronous DE.

Methodology. In the design of original experimental research, the more the extraneous differences between treatment and control can be minimized, the stronger the causal assertion. However, in a meta-analysis, actual control cannot be applied to the studies under scrutiny, so the best that can be done is to estimate the methodological strength or weakness of the research literature. The first thing we found is that methodology is a good predictor of achievement and attitude effect sizes, but a better predictor in synchronous DE studies (49% and 47%, respectively) than in asynchronous DE studies (12% and 22%). Second, we found that methodology is a strong predictor of achievement and attitude effect size, whether entered on the first or the third step of WMR for synchronous DE, but not for asynchronous DE. Because of the way methodology was recoded, this means that studies of asynchronous DE are of higher quality than studies of synchronous DE.

Pedagogy and media. Clark (1983, 1994) has argued vociferously that media and technology, used in educational practice, have no effect on learning. Instead, it is the characteristics of instructional design, such as the instructional strategies that are used, the feedback that is provided, and the degree of learner engagement, that create the conditions within

Bernard et al.

which purposive learning will occur. In general, we found this to be the case. Characteristics of pedagogy tended to take precedence over media, no matter on which step in WMR they were entered. This is especially true for achievement outcomes; the relationship for attitudes is a little more complex. Does this mean that media aren't important? No, it can't mean that because media are a requirement for DE to exist in the first place. It does mean, however, that instructional practices, independent of the medium, are critical to all forms of educational practice, including, and perhaps especially, DE. This seems almost too axiomatic to state, and yet in the literature of DE there is an exaggerated emphasis on the medium *du jour*. As Richard Clark recently explained (personal communication, April and October, 2003), it was the tendency of educational technologists to become enamored with "the toys of technology" that led to his original thesis and his continued insistence that media are of little concern compared with the myriad elements of sound instructional practice. There is a now old instructional design adage that goes something like this: "a medium should be selected in the service of instructional practices, not the other way around." We would encourage all practitioners and policymakers, bent on developing and delivering quality DE, whether on the Internet or through synchronous teleconferencing, to heed this advice.

Considerations for Practice

Before moving on to a discussion of individual study features, there are two issues that need reiteration. First, interpretation of individual predictors in WMR, when overall results are heterogeneous, must proceed with caution (Hedges & Olkin, 1996). Second, some of the individual study feature results are based on a fairly small number of actual outcomes and therefore must be taken as speculative.

Specific considerations. Unfortunately, we are unable to offer any recipes for the design and development of quality DE. Missing information in the research literature, we suspect, is largely responsible for this. However, we are able to speak in broad terms about some of the things that matter in synchronous and asynchronous DE applications:

- Attention to quality course design should take precedence over attention to the characteristics of media. This presumably includes what the instructor does as well as what the student does, although we see only limited direct evidence of either. However, the appearance of "use of systematic instructional design" as a predictor of attitude outcomes implicates instructors and designers of asynchronous DE conditions.
- Active learning (e.g., PBL) that includes (or induces) some collaboration among students appears to foster better achievement and attitude outcomes in asynchronous DE.
- Opportunities for communication, both face-to-face and through mediation, appear to benefit students in synchronous and asynchronous DE.
- "Supplementary one-way video materials" and "use of computer-based instruction" were also found to help promote better achievement and attitude outcomes in synchronous and asynchronous DE.
- In asynchronous DE, media that support interactivity (i.e., CMC and the telephone) appear to facilitate better attitudes, and "providing advance course information" benefits achievement outcomes.

The similarities between the results of achievement and attitude across synchronous and asynchronous DE are both strikingly similar and strikingly different. For instance, for asynchronous DE, problem-based learning (PBL) appears as a strong predictor in favor of the

DE condition. Although this is one of the study features with relatively few instances, we speculate that it is the collaborative, learner-oriented aspect of this instructional strategy that accounts for better achievement and more positive attitudes. Judging from reviews in the medical education literature (e.g., Albanese & Mitchell, 1993; Colliver, 1999), where 30 years of studies have been performed with PBL, this instructional strategy represents a useful mechanism for engaging students, teaching problem-solving, and developing collaborative working skills. Bernard, Rojo de Rubalcava, & St. Pierre (2000) describe ways that PBL might be linked to collaborative learning in online learning environments.

Among the other pedagogical study features is a group of features that relate to both face-to-face and mediated contact with the instructor in a course and among student peers. We also found that “encouragement of contact (either face-to-face or mediated)” predicted outcomes for both synchronous and asynchronous DE, when achievement and attitudes were examined jointly. This suggests that DE should not be a solitary experience as it often was in the era of correspondence education. Instructionally relevant contact with instructors and peers is not only desirable, it is probably necessary for creating learning environments that lead to desirable achievement gains and general satisfaction with DE. This is not a particular revelation, but it is an important aspect of quality course design that should not be neglected or compromised.

One of the surprising aspects of this analysis is that the mechanisms of mediated communication (e.g., e-mail) did not figure more prominently as predictors of learning or attitude outcomes. Computer-mediated communication did arise as a significant predictor of attitude outcomes, but a rather traditional medium, the telephone, also contributed to the media equation. In addition, non-interactive one-way TV/video rose to the top as a significant predictor. However, the results for achievement and attitude were exactly the reverse of each other in this regard. For achievement, TV/video improved DE conditions for both synchronous and asynchronous DE, while use of the telephone favored classroom conditions in synchronous DE. For attitudes, TV/video favored the classroom and use of the telephone favored DE, both in synchronous and asynchronous DE settings. Generally speaking, these results appear to further implicate communication and the use of supplementary visual materials.

If one over-arching generalization is applicable here, it is that sufficient opportunities for both student/instructor and student/student communication are important, possibly in the service of collaborative learning experiences such as problem-based learning. We encourage practitioners to build more of these two elements into DE courses and into classroom experiences as well. We also favor an interpretation of media features as aids to these seemingly important instructional/ pedagogical aspects of course design and delivery. For DE, in particular, where media are the only means of providing collaborative and communicative experiences for students, we see pedagogy and the media that support it working in tandem and not as competing entities in the course developer/instructors set of tools. So, while we have attempted to separate pedagogy from media to assess their relative importance, it is the total package in DE that must ultimately come together to foster student learning and satisfaction.

General considerations. Researchers, educators, and the business community have all commented recently on the future of education and the goals of schooling. These comments focus on the importance of encouraging learners to have a lifelong commitment to learning, to be responsible for their own learning, to have effective interpersonal and communication skills, to be aware of technology as a tool for learning, and to be effective problem solvers with skills transferable to varied contexts. These comments also recognize that learners who have genuine learning goals are likely to remain personally committed to their achievement goals, use complex cognitive skills, and draw upon the active support of the learning community to enhance their

personal skills. These concerns apply with equal if not greater force to learning at a distance, where the challenges of isolation may exacerbate them.

The results of this meta-analysis provide general support for the claim that effective DE depends on the provision of pedagogical excellence. How is this achieved in a DE environment? Particular predictors of pedagogical importance included problem-based learning and interactivity, either face-to-face or through mediation, with instructors and other students. Can we make a more general case? We speculate that the keys to pedagogical effectiveness in DE center on the appropriate and strategic use of interactivity among learners and with the learning material leading to learner engagement, deep processing, and understanding. By what means might interactivity occur?

First, interactivity among learners occurs when technology is used as a communication device and learners are provided with appropriate collaborative activities and strategies for learning together. Here we distinguish between “surface” interaction among learners, where superficial learning is promoted through efficient communication (e.g., seeking only the correct answer), and “deep” interaction among learners, where complex learning is promoted through effective communication (e.g., seeking an explanation). The teacher plays roles here by participating in establishing, maintaining, and guiding interactive communication.

Second, the design of interactivity around learning materials might focus on notions derived from cognitive psychology, including socio-cognitive and constructivist principles of learning such as those summarized by the American Psychological Association (1995, 1997). Additionally, learning materials and tasks must engage the learner in ways that promote meaningfulness, understanding, and transfer. Clarity, expressiveness, and feedback may help to insure learner engagement and interactivity; multimedia-learning materials may do likewise when linked to authentic learning activities.

Considerations for Policymakers

One possible implication is that DE needs to exploit media in ways that take advantage of its power; not just DE as an electronic copy of paper-based material. This may explain why the effect sizes are so small in the current meta-analysis. That is, there is a widespread weakness in the tools of DE. Where are the cognitive tools that encourage deeper, active learning—the ones that Kozma and Cobb predicted would transform learning experience? These need further development and more appropriate deployment. A contrasting view, supported by the size of effects encountered in this quantitative review, is that DE effectiveness is most directly affected by pedagogical excellence rather than media sophistication or flexibility.

The first alternative is a longstanding speculation that likely may not be verified until the next generation of DE is widely available and appropriately used. The second alternative requires that policymakers devote energies to ensuring that excellence and effectiveness take precedence over cost efficiency.

Considerations for Future DE Research

What does this analysis suggest about future DE research directions? The answer to this question depends, to some extent, upon whether we accept the premise of Clark and others that media comparison studies (and DE comparison studies, by extension) answer few useful questions, or the premise of Smith and Dillon (1999) that there is still a place for comparative studies, performed under certain conditions. It is probably true that, once DE is established as a “legitimate alternative to classroom instruction,” the need for comparative DE studies diminishes. After all, even in the world of folklore, the comparison between a steam-driven device and the brawn of John Henry was performed only once, to the demise of John. But it is

also true that before we forge ahead into an indeterminate future, possibly embracing untested fads and following false leads, while at the same time dismantling the infrastructure of the past, we should reflect upon why we are going there and what we risk if we are wrong. And if there is a practical way of translating what we know about “best practices in the classroom” to “best practices in cyberspace,” then a case for continued research in both venues, simultaneously, might be made.

So what can we learn from classroom instruction that can be translated into effective DE practices? One of the few significant findings that emerged from the TV studies of the 50s and 60s was that planning and design pay off—it was not the medium that mattered so much as what came before the TV cameras were turned on. Similarly, in this millennium, we might ask if there are aspects of design, relating to either medium or method, which are optimal in either or both instructional contexts. In collecting these studies we found few factorial designs, suggesting that the bulk of the studies asked questions in the form of, “Is it this or that?” Comparisons such as this are the stock-in-trade of meta-analysis, but once the basic question is answered, more or less, we should begin to move towards answering more subtle and sophisticated questions. More complex designs might enable us to address questions such as “What does it depend on or what moderates between *this* and *that*?” Simply knowing that something works or doesn’t work without knowing why strands us in a quagmire of uncertainty allowing the “gimmick of the week” to become king. It is the examination of the details of research studies that can tell us the “why.”

So if comparison studies do continue—and we suspect that they will—can we envisage an optimal comparative study? In the best of all Campbell and Stanley (1963) worlds, an experiment that intends to establish cause eliminates all rival hypotheses and varies only one aspect of the design—the treatment. Here, it means eliminating all potential confounds—selection, history, materials, etc.—except distance, the one feature that distinguishes *distance* education from face-to-face instruction. The problem is that even if exactly the same media are used in both the DE and the classroom conditions, they are used for fundamentally different purposes, in DE to bridge the distance gap (e.g., online collaborative learning instead of face-to-face collaboration) and in the classroom as a supplement to face-to-face instruction. So, without even examining the problem of media/method confounds and other sources of inequality between treatments, we have already identified a fundamental stumbling block to deriving any more useful information from comparative studies. This does not mean, of course, that imperfectly designed but perfectly described studies (i.e., descriptions of the details of treatments and methodology) are not useful in the hands of a meta-analyst, but will we learn any more than we already know by continuing to pursue comparative research? We suspect not, unless such studies are designed to assess the “active ingredients” in each application, as suggested by Smith and Dillon.

So, what is the alternative? In the realm of synchronous DE, a productive set of studies might involve two classroom/DE dyads, run simultaneously, with one of a host of instructional features being varied across the treatments. In a study of this sort, media are used for the *same* purpose in both conditions, and so *distance is not* the variable under study. In asynchronous DE, we envisage similar direct comparisons between equivalent DE treatments. Bernard and Naidu (1992) performed a study of this sort comparing different conditions of concept mapping and questioning among roughly equivalent DE groups. Studies such as this could even examine different types of media or media used for different purposes without succumbing to the fatal flaw that is inherent in DE/classroom-comparative research.

Here are some other directions for future research:

Bernard et al.

- Developing a theoretical framework for the design and analysis of DE. Adapting the learner-centered principles of the American Psychological Association (1995,1997; see also Lambert & McCombs, 1998) may be a starting point for exploring the cognitive and motivational processes involved in learning at a distance.
- Exploring more fully student motivational dispositions in DE, including task choice, persistence, mental effort, efficacy, and perceived task value. Interest/satisfaction may not indicate success, but the opposite, since students may spend less effort learning, especially when they choose between DE and regular courses for convenience purposes (i.e., happy to have choice and satisfied, but because they may wish to make less of an effort to learn, they are merely conveniencing themselves).
- Examining new aspects of pedagogical effectiveness and efficiency, including faculty development and teaching time, student access and learning time, and cost effectiveness (e.g., cost per student). Establishing desirable skill-sets for instructors of synchronous and asynchronous DE settings might be a place to start. Examining different methods for developing these skill-sets might extend from this examination.
- Studying levels of learning (e.g., simple knowledge or comprehension vs. higher-order thinking). Examining various instructional strategies for achieving these outcomes, such as PBL and collaborative online learning, could represent a very productive line of inquiry.
- Examining inclusivity and accessibility for home learners, rural and remote learners, and learners with various disabilities. Here in particular the appropriate comparison may be with “no instruction” rather than “traditional” classroom instruction.
- Using more rigorous and complete research methodologies, including more detailed descriptions of control conditions in terms of both pedagogical features and media characteristics.

There is one thing that is certain. The demand for research will always lag behind the supply of research, and for this very reason, it is important to apportion our collective research resources judiciously. It may just be that at this point in our evolution, and with so many pressing issues to examine as Internet applications of DE proliferate, continuing to compare DE with the classroom, without attempting to answer the attendant concerns of “why” and “under what conditions,” is wasted time and effort.

Conclusion

This meta-analysis represents a rigorously applied examination of the comparative literature of DE with regard to the variety of conditions of study features and outcomes that are publicly available. We found evidence, in an overall sense, that classroom instruction and DE are comparable, as have some others. However, the wide variability present in all measures precludes any firm declarations of this sort. We confirm the prevailing view that, in general, DE research is of low quality, particularly in terms of internal validity (i.e., control for confounds and inequalities). We found a dearth of information in the literature; a more replete literature could have led to stronger conclusions and recommendations for practice and policymaking. Beyond that, we have also contributed the following: a) a view of the differences that exist in all measures between synchronous and asynchronous DE; b) a view of the relationship between pedagogy and media which appears to be a focus for debate whenever a new learning orientation (e.g., constructivism) or medium of instruction (e.g., computer-mediated communication) appears on the educational horizon; c) an assessment of the relative strength and effect of

methodological quality on the assessment of other contributing factors; d) a glimpse of the relatively few individual study features that predict learning and attitude outcomes; and e) a view of the heterogeneity in findings that hampered our attempts to form homogeneous subsets of study features that could have helped to establish what makes DE better or worse than classroom instruction.

References

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

- Abrami, P. C., & Bures, E. M. (1996). Computer-supported collaborative learning and distance education. *American Journal of Distance Education, 10*(2), 37-42.
- Abrami, P. C., Cohen, P., & d'Apollonia, S. (1988). Implementation problems in meta-analysis. *Review of Educational Research, 58*(2), 151-179.
- Albanese, M. A. & Mitchell, S. (1993). Problem-based learning: A review of literature on its outcomes and implementation issues. *Academic Medicine, 68*(1), 52-81.
- Allen, M., Bourhis, J., Burrell, N., & Mabry, E. (2002). Comparing student satisfaction with distance education to traditional classrooms in higher education: A meta-analysis. *The American Journal of Distance Education, 16*(2), 83-97.
- American Psychological Association (Division 15, Committee on Learner-Centered Teacher Education for the 21st Century). (1995, 1997). *Learner-centered psychological principles: Guidelines for teaching educational psychology in teacher education programs*. Washington, DC: American Psychological Association.
- *Anderson, M. R. (1993). *Success in distance education courses versus traditional classroom education courses*. Unpublished doctoral dissertation, Oregon State University, Corvallis.
- Anglin, G., & Morrison, G. (2000). An analysis of distance education research: Implications for the instructional technologist. *Quarterly Review of Distance Education, 1*(3), 180-194.
- *Appleton, A. S., Dekkers, J., & Sharma, R. (1989, August). *Improved teaching excellence by using Tutored Video Instruction: An Australian case study*. Paper presented at the 11th EAIR Forum. Trier, Germany.
- *Armstrong-Stassen, M., Landstorm, M., & Lumpkin, R. (1998). Student's reactions to the introduction of videoconferencing for classroom instruction. *The Information Society, 14*, 153-164.
- *Bacon, S. F., & Jakovich, J. A. (2001). Instructional television versus traditional teaching of an introductory psychology course. *Teaching of Psychology, 28*(2), 88-91.
- *Bader, M. B., & Roy, S. (1999). Using technology to enhance relationships in interactive television classrooms. *Journal of Education for Business, 74*(6), 357-362.
- *Barber, W., Clark, H., & McIntyre, E. (2002). Verifying success in distance education. *Proceedings of the World Conference on E-Learning in Corp., Govt., Health, & Higher Ed. 2002* (1), 104-109.
- *Barker, B. M. (1994). *Collegiate aviation review. September 1994*. Auburn, AL: University Aviation Association.
- *Barkhi, R., Jacob, V. S., & Pirkul, H. (1999). An experimental analysis of face-to-face versus computer mediated communication channels. *Group Decision and Negotiation, 8*, 325-347.
- *Barnett-Queen, T., & Zhu, E. (September, 1999). *Distance education: Analysis of learning preferences in two sections of undergraduate HBSE-like human growth and development course: Face-to-face and web-based distance learning*. Paper presented at the 3rd Annual

Bernard et al.

Technology Conference for Social Work Education and Practice, Charleston, SC.

- *Bartel, K. B. (1998). *A comparison of students taught utilizing distance education and traditional education environments in beginning microcomputer applications classes at Utah State University*. Unpublished doctoral dissertation, Utah State University, Logan.
- Bates, A. W. (1997). The future of educational technology. *Learning Quarterly*, 2(1), 7-16.
- *Bauer, J. W., & Rezabek, L. L. (1993, September). *Effects of two-way visual contact on verbal interaction during face-to-face and teleconferenced instruction*. In Art, Science & Visual Literacy: Selected Readings from the Annual Conference of the International Visual Literacy Association . Pittsburgh, PA.
- *Beare, P. L. (1989). The comparative effectiveness of videotape, audiotape, and tele-lecture in delivering continuing teacher education. *The American Journal of Distance Education*, 3(2), 57-66.
- Beaudoin, M. (1990). The instructors changing role in distance education. *The American Journal of Distance Education*, 4(2), 21-29.
- *Benbunan, R. (1997). *Effects of computer-mediated communication systems on learning, performance and satisfaction: A comparison of groups and individuals solving ethical scenarios*. Unpublished doctoral dissertation, The State University of New Jersey, Newark, NJ.
- *Benbunan-Fich, R., Hiltz, S. R., & Turoff, M. (2001). A comparative content analysis of face-to-face vs. ALN-mediated teamwork. *Proceedings of the 34th Hawaii International Conference on System Sciences — 2001*.
- Berge, Z. L., & Mrozowski, S. (2001). Review of research in distance education, 1990 to 1999. *American Journal of Distance Education*, 15(3), 15-19.
- Bernard, R. M., & Naidu, S. (1990). Integrating research into practice: The use and abuse of meta-analysis. *Canadian Journal of Educational Communication*, 19(3), 171-195.
- Bernard, R. M., & Naidu, S. (1992). Concept mapping, post-questioning and feedback: A distance education field experiment. *British Journal of Educational Technology*, 23(1), 48-60.
- Bernard, R. M., Rojo de Rubalcava, B., & St. Pierre, D.. (2000). Collaborative online distance education: Issues for Future Practice and Research. *Distance Education*, 21(2), 260-277.
- *Bischoff, W. R., Bisconer, S. W., Kooker, B. M., & Woods, L.C. (1996). Transactional distance and interactive television in the distance education of health professionals. *The American Journal of Distance Education*, 10(3), 4-19.
- *Bisciglia, M. G., & Monk-Turner, E. (2002). Differences in attitudes between on-site and distance-site students in group teleconference courses. *The American Journal of Distance Education*, 16(1), 37-52.
- *Boulet, M. M., Boudreault, S., & Guerette, L. (1998). Effects of a television distance education course in computer science. *British Journal of Educational Technology*, 29(2), 101-111.
- *Britton, O. L. (1992). *Interactive distance education in higher education and the impact of delivery styles on student perceptions*. Unpublished doctoral dissertation, Wayne State University, Detroit, MI.
- *Brown, B. W., & Liedholm, C. E. (2002). Can web courses replace the classroom in principles of microeconomics? *The American Economic Review*, 92(2), 444-448.
- *Browning, J. B. (1999). *Analysis of concepts and skills acquisition differences between web-delivered and classroom-delivered undergraduate instructional technology courses*. Unpublished doctoral dissertation, North Carolina State University, Raleigh, NC.
- *Bruning, R., Landis, M., Hoffman, E., & Grosskopf, K. (1993). Perspectives on an interactive satellite-based Japanese language course. *The American Journal of Distance Education*,

Bernard et al.

7(3), 22-38.

- *Buchanan, E., Xie, H., Brown, M., & Wolfram, D. (2001). A systematic study of web-based and traditional instruction in an MLIS program: Success factors and implications for curriculum design. *Journal of Education for Library and Information Science*, 42(4), 274-288.
- *Burkman, T.A. (1994). *An analysis of the relationship of achievement, attitude, and sociological element of individual learning style of students in an interactive television course*. Unpublished doctoral dissertation, Western Michigan University, Kalamazoo, MI.
- *Cahill, D., & Catanzaro, D. (1997). Teaching first-year Spanish on-line. *Calico Journal*, 14(2-4), 97-114.
- *Callahan, A. L., Givens, P. E., & Bly, R. (1998, June). *Distance education moves into the 21st century: A comparison of delivery methods*. Paper presented at the American Society for Engineering Education Annual Conference, Seattle, WA.
- *Campbell, M., Floyd, J., & Sheridan, J. B. (2002). Assessment of student performance and attitudes for courses taught online versus onsite. *The Journal of Applied Business Research*, 18(2), 45-51.
- Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. New York: Houghton Mifflin.
- *Card, K. A., & Horton, L. (1998, November). *Fostering collaborative learning among students taking higher education administrative courses using computer-mediated communication*. *ASHE annual meeting paper*. Paper presented at the Annual Meeting of the Association for the Study of Higher Education, Miami, FL.
- *Carey, J. M. (2001). *Effective student outcomes: A comparison of online and face-to-face delivery modes*. Retrieved April 30, 2003 from http://teleducation.nb.ca/content/pdf/english/DEOSNEWS_11.9_effective-student-outcomes.pdf
- *Carl, D. R., & Densmore, B. (1988). Introductory accounting on distance university education via television (duet): A comparative evaluation. *Canadian Journal of Educational Communication*, 17(2), 81-94.
- Carpenter, C. R., Greenhill, L. P. (1955). *An Investigation of Closed-Circuit Television for Teaching University Courses*. Instructional Television Research, Project Number One. The Pennsylvania State University.
- Carpenter, C. R., & Greenhill, L. P. (1958). *An Investigation of Closed-Circuit Television for Teaching University Courses*. Instructional Television Research, Report Number Two. The Pennsylvania State University.
- *Carrell, L. J., & Menzel, K. E. (2001). Variations in learning, motivation, and perceived immediacy between live and distance education classrooms. *Communication Education*, 50(3), 230-240.
- *Casanova, R. S. (2001). *Student performance in an online general college chemistry course*. Retrieved April 30, 2003 from <http://www.chem.vt.edu/confchem/2001/c/04/capefear.html>
- *Caulfield, J. L. (2001). *Examining the effect of teaching method and learning style on work performance for practicing home care clinicians*. Unpublished doctoral dissertation, Marquette University, Milwaukee, WI.
- 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. Retrieved April 30, 2001 from

Bernard et al.

<http://www.unf.edu/~ccavanau/CavanaughIJET01.pdf>

- *Chapman, A. D. (1996). *Using interactive video to teach learning theory to undergraduates: Problems and benefits*. (ERIC Document Reproduction Service ED 406 425)
- *Chen, I. M. C. (1991). *The comparative effectiveness of satellite and face-to-face delivery for a short-term substance abuse education program*. Unpublished doctoral dissertation, University of Missouri, Kansas City.
- *Cheng, H. C., Lehman, & Armstrong, P. (1991). Comparison of performance and attitude in traditional and computer conferencing classes. *The American Journal of Distance Education*, 5(3), 51-64.
- Chickering, A. & Gamson, Z. (1987). Seven principles of good practice in undergraduate education. *AAHE Bulletin*, 39(2), 3-7.
- *Cho, E. (1998). Analysis of teacher & students' attitudes on two-way video tele-educational system for Korean elementary school. *Educational Technology Research and Development*, 46(1), 98-105.
- *Chung, J. (1991). Televised teaching effectiveness: Two case studies. *Educational Technology*, 31(1), 41-47.
- *Chute, A. G., Balthazar, L. B., & Posten, C. O. (1988). Learning from tele-training. *The American Journal of Distance Education*, 2(3), 55-63.
- *Chute, A. G., Hulik, M., & Palmer, C. (1987, May). *Tele-training productivity at AT&T*. Paper presented at International Teleconferencing Association Annual Convention, Washington, DC.
- *Cifuentes, L., & Hughey, J. (1998, February). Computer conferencing and multiple intelligences: Effects on expository writing. *Proceedings of selected research and development presentations at the National Convention of the Association for Educational Communications and Technology (AECT)*. St. Louis, MO.
- *Clack, D., Talbert, L., Jones, P., & Dixon, S. (2002). *Collegiate skills versatile schedule courses*. Retrieved April 30, 2003 from <http://www.schoolcraft.cc.mi.us/leagueproject/pdfs/documents/Learning%20First%20Winter%202002.pdf>
- *Clark, B. A. (1989, March 12). *Comparisons of achievement of students in on-campus classroom instruction versus satellite teleconference instruction*. Paper presented at the National Conference on Teaching Public Administration, Charlottesville, VA.
- Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*. 53(4), 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(1), 3-16.
- *Coates, D., Humphreys, B. R., Kane, J., Vachris, M., Agarwal, R., & Day, E. (2001, January). *"No significant distance" between face to face and online instruction: Evidence from principles of economics*. Paper presented at the Meeting of the Allied Social Science Association, New Orleans, Louisiana.
- Cobb, T. (1997). Cognitive efficiency: Toward a revised theory of media. *Educational Technology Research & Development*, 45(4), 21-35.
- *Coe, J. A. R., & Elliott, D. (1999). An evaluation of teaching direct practice courses in a distance education program for rural settings. *Journal of Social Work Education*, 35(3), 353-365.
- *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.
- Colliver, J. (1999). Effectiveness of PBL curricula. *Medical Education* 34(11), 959-960.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin Company.
- *Cooper, L. W. (2001). A comparison of online and traditional computer applications classes. *T.H.E. Journal*, 28(8), 52-58.
- *Cordover, P. P. (1996). *A comparison of a distance education and locally based course in an urban university setting*. Unpublished doctoral dissertation, Florida International University, Miami, FL.
- *Croson, R. T. A. (1999). Look at me when you say that: An electronic negotiation simulation. *Simulation & Gaming*, 30(1), 23-37.
- *Cross, R. F. (1996). Video-taped lectures for honours students on international industry based learning. *Distance Education*, 17(2), 369-386.
- *Curnow, C. K. (2001). *Social interaction, learning styles, and training outcomes: Differences between distance learning and traditional training*. Unpublished doctoral dissertation, The George Washington University, Washington, DC.
- *Dalton, B. (1999). *Evaluating distance education*. Retrieved April 17, 2003 from www.sc.edu/cosw/PDFs/daltonb.pdf
- *Davis, J. D., Odell, M., Abbitt, J., & Amos, D. (1999). Developing online courses: A comparison of Web-based instruction with traditional instruction. *Proceedings of SITE — Society for Information Technology & Teacher Education International Conference*.
- *Davis, J. L. (1996). Computer-assisted distance learning, part II: Examination performance of students on & off campus. *Journal of Engineering Education*, (January issue), 77-82.
- *Davis, R. S., & Mendenhall, R. (1998). *Evaluation comparison of online and classroom instruction for HEPE 129 — Fitness and Lifestyle Management Course*. Brigham Young University, (ERIC Document Reproduction Service No. ED 427-752)
- *Day, T. M., Raven, M. R., & Newman, M. E. (1998). The effects of world wide web instruction and traditional instruction and learning styles on achievement and changes in student attitudes in a technical writing in agricomunication course. *Journal of Agricultural Education*, 39(4), 65-75.
- Dede, C. (1996). The evolution of distance education: Emerging technologies and distributed learning. *The American Journal of Distance Education*, 10(2), 4-36.
- *Dees, S. C. (1994). *An investigation of distance education versus traditional course delivery using comparisons of student achievement scores in advanced placement chemistry and perceptions of teachers and students about their delivery system*. Unpublished doctoral dissertation, Northern Illinois University, DeKalb.
- *Dexter, D. J. (1995). *Student performance-based outcomes of televised interactive community college*. Unpublished doctoral dissertation, Colorado State University, Fort Collins.
- *Diaz, D. P. (2000). *Comparison of student characteristics, and evaluation of student success, in an online health education course*. Unpublished doctoral dissertation, Nova Southeastern University, Fort Lauderdale, FL.
- Diaz, D. P. (2000, March/April). Carving a new path for distance education research. *The Technology Source*. Retrieved July 24, 2001 from <http://horizon.unc.edu/TS/default.asp?show=articleandid=68>
- *DiBartola, L. M., Miller, M. K., & Turley, C. L. (2001). Do learning style and learning environment affect learning outcome? *Journal of Allied Health*, 30(2), 112-115.
- *Dillon, C. L., Gunawardena, C. N., & Parker, R. (1992). Learner support: The critical link in distance education. *Distance Education*, 13(1), 29-45.

Bernard et al.

- Dillon, C. L., & Walsh, S. M. (1992). Faculty: The neglected resource in distance education. *The American Journal of Distance Education*, 6(3), 5-21.
- *Dominguez, P. S., & Ridley, D. R. (2001). Assessing distance education courses and discipline differences in their effectiveness. *Journal of Instructional Psychology*, 28(1), 15-19.
- *Dutton, J., Dutton, M., & Perry, J. (2001). *Do online students perform as well as lecture students?* Retrieved April 28, 2003 from <http://www4.ncsu.edu/unity/users/d/dutton/public/research/online.pdf>
- *Egan, M. W., McCleary, I. D., Sebastian, J. P., & Lacy, H. (1988). Rural preservice teacher preparation using two-way interactive television. *Rural Special Education Quarterly*, 9(3), 27-37.
- *Fallah, H. M., & Ubell, R. (2000). *Blind scores in a graduate test: Conventional compared with web-based outcomes.* Retrieved April 17, 2003 from www.aln.org/publications/magazine/v4n2/fallah.asp
- *Faux, T. L., & Black-Hughes, C. (2000). A comparison of using the Internet versus lectures to teach social work history. *Research on Social Work Practice*, 10(4), 454-466.
- *Flaskerud, G. (1994). *The effectiveness of an interactive video network (INV) extension workshop.* Pennsylvania State University; University Park, PA.
- *Flowers, C., Jordan, L., Algozzine, B., Spooner, F., & Fisher, A. (2001). Comparison of student rating of instruction in distance education and traditional courses. *Proceedings of the SITE — Society for Information Technology & Teacher Education International Conference 2001* (1), 2314-2319.
- *Foell, N. A. (1989, December). *Using computers to provide distance learning, the new technology.* Paper presented at the Annual Meeting of the American Vocational Education Research Association, Orlando, FL.
- *Freitas, F. A., Myers, S. A., & Avtgis, T. A. (1998). Student perceptions of instructor immediacy in conventional and distributed learning classrooms. *Communication Education*, 47(4), 366-372.
- *Fritz, S., Bek, T. J., & Hall, D. L. (2001). Comparison of campus and distance undergraduate leadership students' attitudes. *The Journal of Behavioral and Applied Management*, 3(1), 3-12.
- *Fulmer, J., Hazzard, M., Jones, S., & Keene, K. (1992). Distance learning: An innovative approach to nursing education. *Journal of Professional Nursing*, 8(5), 289-294.
- *Furste-Bowe, J. A. (1997). Comparison of student reactions in traditional and videoconferencing courses in training and development. *International Journal of Instructional Media*, 24(3), 197-205.
- *Fyock, J. J. (1994). *Effectiveness of distance learning in three rural schools as perceived by students (student-rated).* Unpublished doctoral dissertation, Cornell University, Ithaca, NY.
- Garrison, D. R., Anderson, T. & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, 15(1), 7-23.
- Garrison, D. R., & Shale, D. (1987). Mapping the boundaries of distance education: Problems in defining the field. *The American Journal of Distance Education*, 1(1), 4-13.
- *Gee, D. D. (1991). *The effects of preferred learning style variables on student motivation, academic achievement, and course completion rates in distance education.* Unpublished doctoral dissertation, Texas Tech University, Lubbock.
- Gehlauf, D. N., Shatz, M. A., & Frye, T. W. (1991). Faculty perceptions of interactive television instructional strategies: Implications for training. *The American Journal of Distance*

Bernard et al.

Education, 5(3), 20-28.

- *George Mason University, Office of Institutional Assessment (2001). *Technology in the curriculum: An assessment of the impact of on-line courses*. Retrieved April 17, 2003 from <http://assessment.gmu.edu/reports/Eng302/Eng302Report.pdf>
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- *Glenn, A. S. (2001). *A comparison of distance learning and traditional learning environments*. (ERIC Document Reproduction Service No. ED 457 778)
- *Goodyear, J. M. (1995). *A comparison of adult students' grades in traditional and distance education courses*. Unpublished doctoral dissertation, University of Alaska Anchorage, Anchorage.
- Graham, C., Cagiltay, K., Craner, J., Lim, B., & Duff, T.M. (2000). *Teaching in a Web-based distance learning environment: An evaluation summary based on four courses*. Center for Research on Learning and Technology Technical Report No. 13-00. Indiana University, Bloomington.
- *Gray, B. A. (1996). *Student achievement and temperament types in traditional and distance learning environments (urban education, traditional education)*. Unpublished doctoral dissertation, Wayne State University, Detroit, MI.
- *Grayson, J. P., MacDonald, S. E., & Saindon, J. (2001). *The efficacy of web-based instruction at York University: A case study of Modes of Reasoning, 1730*. Retrieved May 13, 2003 from <http://www.atkinson.yorku.ca/~pgrayson/areport1.pdf>
- *Grimes, P. W., Neilsen, J. E., & Ness, J. F. (1988). The performance of nonresident students in the "economics USA" tele-course. *The American Journal of Distance Education*, 2(2), 36-43.
- *Hackman, M. Z., & Walker, K. B. (1994, July). *Perceptions of proximate and distant learners enrolled in university-level communication courses: a significant non-significant finding*. Paper presented at the 44th Annual Meeting of the International Communication Association, Sydney, New South Wales, Australia.
- *Hahn, H. A., Ashworth, R.L., Phelps, R.H., Wells, R.A., Richards, R.E., Daveline, K.A. (1991). *Distributed training for the reserve component: Remote delivery using asynchronous computer conferencing* (Report No. 1581). Idaho, Falls: U.S. Army Research Institute for the Behavioral and Social Sciences. (ERIC Document Reproduction Service No. ED 359 918)
- *Harrington, D. (1999). Teaching statistics: A comparison of traditional classroom and programmed instruction/distance learning approaches. *Journal of Social Work Education*, 35(3), 343-352.
- *Hassenplug, C. A., Karlin, S., & Harnish, D. (1995). *A statewide study of factors related to the successful implementation of GSAMS credit courses at technical institutes*. Athens, GA: Occupational Research Group, School of Leadership and Lifelong Learning, College of Education, The University of Georgia. (ERIC Document Reproduction Service No. ED 391 891)
- Hawkes, M. (2001). Variables of interest in exploring the reflective outcomes of network-based communication. *Journal of Research on Computing in Education*, 33(3) 299-315.
- Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hedges, L. V., Shymansky, J. A., & Woodworth, G. (1989). *A practical guide to modern methods of meta-analysis*. [ERIC Document Reproduction Service No. ED 309 952].
- *Heiens, R. A., & Hulse, D. B. (1996). Two-way interactive television: An emerging technology

Bernard et al.

- for university level business school instruction. *Journal of Education for Business*, 72(2), 74-77.
- *Hilgenberg, C., & Tolone, W. (2001). Student perceptions of satisfaction and opportunities for critical thinking in distance education by interactive video. In M. G. Moore, & J. T. Savrock, (Eds.), *Distance education in the health sciences* University Park, PA: The American Center for the Study of Distance Education.
- *Hiltz, S. R. (1993). Correlates of learning in a virtual classroom. *International Journal of Man Machine Studies*, 39(1), 71-98.
- *Hiltz, S. R. (1997). *Impacts of college-level courses via asynchronous learning networks: Some preliminary results*. Retrieved May 5, 2003 from <http://www.aln.org/publications/jaln/index.asp>
- *Hinnant, E. C. (1994). *Distance learning using digital fiber optics: A study of student achievement and student perception of delivery system quality*. Unpublished doctoral dissertation, Mississippi State University, Starkville.
- *Hittelman, M. (2001). *Distance education report: Fiscal years 1995-1996 through 1999-2000*. Sacramento, CA: California Community Colleges, Office of the Chancellor.
- *Hoban, G., Neu, B., & Castle, S. R. (2002, April). *Assessment of student learning in an educational administration online program*. Paper presented at the American Educational Research Association Annual Meeting, New Orleans, LA.
- *Hodge-Hardin, S. (1997, April). Interactive television vs. a traditional classroom setting: A comparison of student math achievement. *Proceedings of the Mid-South Instructional Technology Conference*. Murfreesboro, TN.
- *Hoey, J. J., Pettitt, J. M., Brawner, C. E., & Mull, S. P. (1998). *Project 25: First semester assessment: A report on the implementation of courses offered on the Internet*. Retrieved March 15, 1997 from http://www2.ncsu.edu/ncsu/ltc/Project25/info/f97_assessment.html
- *Hogan, R. (1997, July). *Analysis of student success in distance learning courses compared to traditional courses*. Paper presented at the 6th Annual Conference on Multimedia in Education and Industry, Chattanooga, TN.
- *Huff, M. T. (2000). A comparison study of live instruction versus interactive television for teaching MSW students' critical thinking skills. *Research on Social Work Practice*, 10(4), 400-416.
- *Hurlburt, R. T. (2001). "Lectlets" delivery content at a distance: Introductory statistics as a case study. *Teaching of Psychology*, 28(1), 15-20.
- *Jeannette, K. J., & Meyer, M. H. (2002). Online learning equals traditional classroom training for master gardeners. *HortTechnology*, 12(1), 148-156.
- *Jenkins, S. J., & Downs, E. (2002). Differential characteristics of students in on-line vs. traditional courses. *Proceedings of SITE — Society for Information Technology & Teacher Education International Conference, 2002*, 194-196.
- *Jewett, F. (1998). *The education network of Maine: A case study in the benefits and costs of instructional television*. Seal Beach, CA: California State University, Seal Beach. Office of the Chancellor.
- *Jewett, F. (1998). *The Westnet program — SUNY Brockport and the SUNY campuses in Western New York state: A case study in the benefits and costs of an interactive television network*. (ERIC Document Reproduction Service No. ED 420 301)
- *Johnson, G. R., O'Connor, M., & Rossing, R. (1985). Interactive two-way television: Revisited. *Journal of Educational Technology Systems*, 13(3), 153-158.
- *Johnson, K. R. (1993). *An analysis of variables associated with student achievement and satisfaction in a university distance education course*. Unpublished doctoral dissertation,

New York State University, Buffalo.

- *Johnson, M. (2002). Introductory biology online: Assessing outcomes of two student populations. *Journal of College Science Teaching*, 31(5), 312-317.
- *Johnson, S. D., Aragon, S. R., Shaik, N., & Palma-Rivas, N. (1999). *Comparative analysis of online vs. face-to-face instruction*. Champaign, IL: Department of Human Resource Education, University of Illinois at Urbana-Champaign.
- *Johnson, S. M. (2001). Teaching introductory international relations in an entirely web-based environment: comparing student performance across and within groups. *Education at a Distance*, 15(10).
- *Jones, E. R. (1999, February). *A comparison of an all web-based class to a traditional class*. Paper presented at the Society for Information Technology & Teacher Education, San Antonio, TX.
- Jung, I., & Rha, I. (2000, July-August). Effectiveness and cost-effectiveness of online education: A review of the literature. *Educational Technology*, 57-60.
- *Kabat, E. J., & Friedel, J. N. (1990). *The Eastern Iowa community college districts televised interactive education evaluation report*. Clinton, IA: Eastern Iowa Community College.
- *Kaeley, G. S. (1989). Instructional variables and mathematics achievement in face-to-face and distance teaching modes. *International Council of Distance Education Bulletin*, 15-31.
- *Kahl, T. N., & Cropley, A. J. (1986). Face-to-face versus distance learning: Psychological consequences and practical implications. *Distance Education*, 7(1), 38-48.
- Kanuka, H., Collett, D., & Caswell, C. (2003). University instructor perceptions of the use of asynchronous text-based discussion in distance courses. *The American Journal of Distance Education*, 16(3), 151-167.
- *Kataoka, H. C. (1987). Long-distance language learning: The second year of televised Japanese. *Journal of Educational Techniques and Technologies*, 20(2), 43-50.
- Keegan, D. (1996). *Foundations of distance education*. (3rd ed). London: Routledge.
- *Keene, S. D., & Cary, J. S. (1990). Effectiveness of distance education approach to U.S. Army Reserve component training. *The American Journal of Distance Education*, 4(2), 14-20.
- Kember, D. (1996). *Open learning courses for adults: A model of student progress*, Englewood Cliffs, NJ: Educational Technology Publications.
- *Kennedy, R. L., Suter, W. N., & Clowers, R. L. (1997, November 12-14). *Research by electronic mail*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Memphis, TN.
- *King, F. B. (2001). *Asynchronous distance education courses employing web-based instruction: Implications of individual study skills self-efficacy and self-regulated learning*. Unpublished doctoral dissertation, University of Connecticut, Storrs.
- *Knox, D. M. (1997). A review of the use of video-conferencing for actuarial education — a three-year case study. *Distance Education*, 18(2), 225-235.
- *Kochman, A., & Maddux, C. D. (2001). Interactive televised distance learning versus on-campus instruction: A comparison of final grades. *Journal of Research on Technology in Education*, 34(1), 87-91.
- Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research & Development*, 42(2), 7-19.
- *Kranc, B. M. (1997). *The impact of individual characteristics on telecommunication distance learning cognitive outcomes in adult/nontraditional students*. Unpublished doctoral dissertation, North Carolina State University, Raleigh, NC.
- *Kretovics, M. A. (1998). *Outcomes assessment: The impact of delivery methodologies and personality preference on student learning outcomes*. Unpublished doctoral dissertation,

- Colorado State University, Fort Collins.
- Lambert, N., & McCombs, B. (1998). Learner-centered schools and classrooms as a direction for school reform. In N. Lambert & B. McCombs (Eds.), *How students learn: Reforming schools through learner-centered education* (pp. 1-22). Washington, DC: American Psychological Association.
- *LaRose, R., Gregg, J., & Eastin, M. (1998). Audiographic telecourses for the Web: An experiment. *Journal of Computer-Mediated Communication [Online]*, 4(2), Retrieved May 15, 2003 from <http://www.ascusc.org/jcmc/vol4/issue2/larose.html#ABSTRACT>
- *Larson, M. R., & Bruning, R. (1996). Participant perceptions of a collaborative satellite-based mathematics course. *The American Journal of Distance Education*, 10(1), 6-22.
- *Lia-Hoagberg, B., Vellenga, B., Miller, M., & Li, T. Y. (1999). A partnership model of distance education: Students' perceptions of connectedness and professionalization. *Journal of Professional Nursing*, 15(2), 116-122.
- *Liang, C. C. (2001). Guidelines for distance education: A case study in Taiwan. *Journal of Computer Assisted Learning*, 17(1), 48-57.
- *Lilja, D. J. (2001). Comparing instructional delivery methods for teaching computer systems performance analysis. *IEEE Transactions on Education*, 44(1), 35-40.
- *Litchfields, R. E., Oakland, M. J., & Anderson, J. A. (2002). Relationships between intern characteristics, computer attitudes, and use of online instruction in a dietetic training program. *The American Journal of Distance Education*, 16(1), 23-36.
- *Logan, E., & Conerly, K. (2002). *Students creating community: An investigation of student interactions in a web-based distance learning environment*. Retrieved April 28, 2003 from www.icte.org/T01_Library/T01_253.pdf
- *Long, L., & Javidi, A. (2001). *A comparison of course outcomes: Online distance learning versus traditional classroom settings*. Retrieved April 28, 2003 from http://www.communication.ilstu.edu/activities/NCA2001/paper_distance_learning.pdf
- Lou, Y. (2004). Learning to solve complex problems through online between-group collaboration. *Distance Education*. 25(1), 50-66.
- Lou, Y., Dedic, H., & Rosenfield, S. (2003). Feedback model and successful e-learning. In S. Naidu (Ed.), *Learning and teaching with technology: Principles and practice* (pp. 249-260). London and Sterling, VA: Kogan Page.
- Lou, Y. & MacGregor, S. K. (2002). *Enhancing online learning with between group collaboration*. Paper presented at the Teaching Online in Higher Education Online Conference, November, 12-14.
- *MacFarland, T. W. (1998). *A comparison of final grades in courses when faculty concurrently taught the same course to campus-based and distance education students: winter term 1997*. Fort Lauderdale, FL: Nova Southeastern University.
- *MacFarland, T. W. (1999). *Matriculation status of fall term 1993 center for psychological studies students by the beginning of fall term 1998: Campus-based students and distance education students by site*. (ERIC Document Reproduction Service No. ED 434 557)
- Machtmes, K., & Asher, J.W. (2000). A meta-analysis of the effectiveness of telecourses in distance education. *American Journal of Distance Education*. 14(1): 27-46.
- *Magiera, F. T. (1994). Teaching managerial finance through compressed video: An alternative for distance education. *Journal of Education for Business*, 69(5), 273-277.
- *Magiera, F. T. (1994-1995). Teaching personal investments via long-distance. *Journal of Educational Technology Systems*, 23(4), 295-307.
- *Maki, R. H., Maki, W. S., Patterson, M., & Whittaker, P. D. (2000). Evaluation of a Web-based introductory psychology course: I. Learning and satisfaction in on-line versus lecture

Bernard et al.

- courses. *Behaviour Research Methods, Instruments & Computers*, 32, 230-239.
- *Maki, W. S., & Maki, R. H. (2002). Multimedia comprehension skill predicts differential outcomes of Web-based and lecture courses. *Journal of Experimental Psychology: Applied*, 8(2), 85-98.
- *Maltby, J. R., & Whittle, J. (2000). *Learning programming online: Student perceptions and performance*. Retrieved April 28, 2003 from www.ascilite.org.au/conferences/coffs00/papers/john_maltby.pdf
- *Martin, E. D., & Rainey, L. (1993). Student achievement and attitude in a satellite-delivered high school science course. *The American Journal of Distance Education*, 7(1), 54-61.
- *Marttunen, M., & Laurinen, L. (2001). Learning of argumentation skills in networked and face-to-face environments. *Instructional Science*, 29 127-153.
- *Maxcy, D. O., & Maxcy, S. J. (1986-1987). Computer/telephone pairing for long distance learning. *Journal of Educational Technology Systems*, 15(2), 201-211.
- *McCleary, I. D., & Egan, M. W. (1989). Program design and evaluation: Two-way interactive television. *The American Journal of Distance Education*, 3(1), 50-60.
- *McGreal, R. (1994). Comparison of the attitudes of learners taking audiographic teleconferencing courses in secondary schools in Northern Ontario. *Interpersonal Computing and Technology Journal*, 2(4), 11-23.
- *McKissack, C. E. (1997). *A comparative study of grade point average (GPA) between the students in traditional classroom setting and the distance learning classroom setting in selected colleges and universities*. Unpublished doctoral dissertation, Tennessee State University, Nashville.
- McKnight, C.B. (2001). Supporting critical thinking in interactive learning environments. *Computers in the Schools*, 17(3-4), 17-32.
- *Mehlenbacher, B., Miller, C., Convington, D., & Larsen, J. (2000). Active and interactive learning online: A comparison of web-based and conventional writing classes. *IEEE Transactions on Professional Communication*, 43(2), 166-184.
- *Miller, J. W., McKenna, M. C., & Ramsey, P. (1993). An evaluation of student content learning and affective perceptions of a two-way interactive video learning experience. *Educational Technology*, 33(6), 51-55.
- *Mills, B. D. (1998). Comparing optimism and pessimism of students in distance-learning and on campus. *Psychological Reports*, 83(3, Pt 2), 1425-1426.
- *Mills, B. D. (1998). Replication of optimism and pessimism of distance-learning and on-campus students. *Psychological Reports*, 83(3, Pt. 2), 1454.
- *Minier, R. W. (2002). *An investigation of student learning and attitudes when instructed via distance in the selection and use of K-12 classroom technology*. Unpublished doctoral dissertation, The University of Toledo, Toledo, OH.
- *Mock, R. L. (2000). *Comparison of online coursework to traditional instruction*. Retrieved April 28, 2003 from <http://hobbes.lite.msu.edu/~robmock/masters/mastersonline.htm#toc>
- *Molidor, C. E. (2000). *The development of successful distance education in social work: A comparison of student satisfaction between traditional and distance education classes*. Retrieved April 23, 2003, from www.nssa.us/nssajrnl/18-1/pdf/14.pdf
- Moore, M., & Thompson, M. (1990). *The effects of distance learning: A summary of the literature*. University Park, PA: The Pennsylvania State University. [ERIC Document Reproduction Service No. ED 391 467].
- *Moorhouse, D. R. (2001). *Effect of instructional delivery method on student achievement in a master's of business administration course at the Wayne Huizenga School of Business and Entrepreneurship*. Ft. Lauderdale, FL: Nova Southeastern University.

Bernard et al.

- Morrison, G. R. (1994). The media effects question: "Unresolvable" or asking the right question. *Educational Technology Research and Development*, 42(2), 41-44.
- *Moshinski, J. F. (1997). The effects of using constructivist learning models when delivering electronic distance education (EDE) courses: A perspective study. *Journal of Instruction Delivery Systems*, 11(1), 14-20.
- Mottet, T. P. (1998). Interactive television instructors = perceptions of students = nonverbal responsiveness and effects on distance teaching. *Dissertation Abstracts International*, 59(02), 460A. (University Microfilms No. AAT98-24007)
- *Murphy, T. H. (2000). An evaluation of a distance education course design for general soils. *Journal of Agricultural Education*, 41(3), 103-113. Retrieved May 15, 2003 from <http://pubs.aged.tamu.edu/jae/pdf/vol41/41-03-103.pdf>
- *Murray, J. D., & Heil, M. (1987). *Project evaluation: 1986-87 Pennsylvania teleteaching project*. Mansfield, PA: Mansfield University of Pennsylvania.
- *Muta, H., Kikuta, R., Hamano, T., & Maesako, T. (1997). The effectiveness of low-cost telelecturing. *Staff and Educational Development International*, 1(2-3), 129-142.
- *Mylona, Z. H. (1999). *Factors affecting enrollment satisfaction and persistence in a Web-based video-based and conventional instruction*. Unpublished doctoral dissertation, University of Southern California, Los Angeles.
- *Nakshabandi, A. A. (1993). A comparative evaluation of a distant education course for female students at King Saud University. *International Journal of Instructional Media*, 20(2), 127-136.
- *Navarro, P., & Shoemaker, J. (1999). *Economics in cyberspace: A comparison study*. Retrieved April 23, 2003 from <http://www.powerofeconomics.com/AJDEFINAL.pdf>
- *Navarro, P., & Shoemaker, J. (1999). *The power of cyberlearning: An empirical test*. Retrieved April 23, 2003 from <http://www.powerofeconomics.com/jchefinalsubmission.pdf>
- *Nesler, M. S., Hanner, M. B., Melburg, V., & McGowan, S. (2001). Professional socialization of baccalaureate nursing students: Can students in distance nursing programs become socialized? *Journal of Nursing Education*, 40(7), 293-302.
- *Neuhauser, C. (2002). Learning style and effectiveness of online and face-to-face instruction. *The American Journal of Distance Education*, 16(2), 99-113.
- *Newlands, D., & McLean, A. (1996). The potential of live teacher supported distance learning: A case-study of the use of audio conferencing at the university of Aberdeen. *Studies in Higher Education*, 21(3), 285-297.
- Nipper, S. (1989). Third generation distance learning and computer conferencing. In: R. Mason and A. Kaye (Eds.), *Mindweave: Communication, computers and distance education* (pp. 63-73). Oxford, UK: Pergamon Press.
- *Obermier, T. R. (1991). *Academic performance of video-based distance education students and on-campus students*. Unpublished doctoral dissertation, Colorado State University, Fort Collins.
- *Ocker, R. J., & Yaverbaum, G. J. (1999). Asynchronous computer-mediated communication versus face-to-face collaboration: Results on student learning, quality and satisfaction. *Group Decision and Negotiation*, 8, 427-440.
- *Olejnik, S., & Wang, L. (1992-1993). An innovative application of the Macintosh Classic II computer for distance education. *Journal of Educational Technology Systems*, 21(2), 87-101.
- Ostendorf, V. A. (1997). Teaching by television. In T. E. Cyr (Ed.), *Teaching and learning at a distance: What it takes to effectively design, deliver, and evaluation programs*. New directions for teaching and learning, No. 71. San Francisco: Jossey-Bass.

Bernard et al.

- *Parker, D., & Gemino, A. (2001). Inside online learning: Comparing conceptual and technique learning performance in place-based and ALN formats. *Journal of Asynchronous Learning Networks [Online]*, 5(2), 64-74. Retrieved May 13, 2003 from http://www.aln.org/publications/jaln/v5n2/v5n2_parkergemino.asp
- *Parkinson, C. F., & Parkinson, S. B. (1989). A comparative study between interactive television and traditional lecture course offerings for nursing students. *Nursing and Health Care*, 10(9), 498-502.
- Perraton, H. (2000). Rethinking the research agenda. *International Review of Research in Open and Distance Learning*, 1(1). Retrieved July 24, 2001 from <http://www.irrodl.org/v1.1.html>
- *Petracchi, H. E., & Patchner, M. A. (2000). Social work students and their learning environment: A comparison of interactive television, face-to-face instruction, and the traditional classroom. *Journal of Social Work Education*, 36(2), 335-346.
- *Petracchi, H. E., & Patchner, M. E. (2001). A comparison of live instruction and interactive televised teaching: A 2-year assessment of teaching an MSW research methods course. *Research on Social Work Practice*, 11(1), 108-117.
- *Phelps, R. H., Wells, R. A., Ashworth, R. L., & Hahn, H. A. (1991). Effectiveness and costs of distance education using computer-mediated communication. *The American Journal of Distance Education*, 5(3), 7-19.
- *Phillips, M. R., & Peters, M. J. (1999). Targeting rural students with distance learning courses: A comparative study of determinant attributes and satisfaction levels. *Journal of Education for Business*, 74(6), 351-356.
- Phipps, R. & Merisotis, J. (1999). *What's the difference? A review of contemporary research on the effectiveness of distance learning in higher education*. Washington, DC: The Institute for Higher Education Policy.
- *Piccoli, G., Ahmad, R., & Ives, B. (2001). Web-based virtual learning environments: A research framework and a preliminary assessment of effectiveness in basic IT skills training. *MIS Quarterly*, 25(4), 401-426.
- *Pirrong, G. D., & Lathen, W. C. (1990). The use of interactive television in business education. *Educational Technology*, 30 (May) 49-54.
- *Pugh, R. C., & Siantz, J. E. (1995, April). *Factors associated with student satisfaction in distance education using slow scan television*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.
- *Reagan, C. (2002). Teaching research methods online: Course development and comparison to traditional delivery. *Proceedings of SITE — Society for Information Technology and Teacher Education International Conference 2002*(1), 141-145.
- *Redding, T. R., & Rotzien, J. (2001). Comparative analysis of online learning versus classroom learning. *Journal of Interactive Instruction Development*, 13(4), 3-12.
- Rekkedal, T., & Qvist-Eriksen, S. (2003). Internet Based E-learning, Pedagogy and Support Systems. Retrieved November 22, 2003 from <http://home.nettskolen.com/~torstein/>
- *Richards, I. E. (1994). *Distance learning: A study of computer modem students in a community college*. Unpublished doctoral dissertation, Kent State University, Kent, OH.
- *Richards, I., & others. (1995, April). *A study of computer-modem students: A call for action*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.
- *Ritchie, H., & Newby, T. J. (1989). Classroom lecture/discussion vs. live televised instruction: A comparison of effects on student performance, attitudes, & interaction. *The American Journal of Distance Education*, 3(3), 36-45.

Bernard et al.

- *Rivera, J., & Rice, M. (2002). A comparison of student outcomes & satisfaction between traditional & web based course offerings. *Online Journal of Distance Learning Administration*, 5(3), Retrieved May 13, 2003 from <http://www.westga.edu/~distance/ojdla/fall53/rivera53.html>
- *Ross, J. L. (2000). *An exploratory analysis of post-secondary student achievement comparing a web-based and a conventional course learning environment*. Unpublished doctoral dissertation, University of Calgary, Calgary, AB, Canada.
- *Rost, R. C. (1997). *A study of the effectiveness of using distance education to present training programs to extension service master gardener trainees*. Unpublished doctoral dissertation, Oregon State University, Corvallis.
- *Ruchti, W. P., & Odell, M. R. L. (2000). *Comparison and evaluation of online and classroom instruction in elementary science teaching methods courses*. Retrieved April 30, 2003 from <http://nova.georgefox.edu/nwcc/arpapers/uidaho.pdf>
- *Rudin, J. P. (1998). Teaching undergraduate business management courses on campus and in prisons. *Journal of Correctional Education*, 49(3), 100-106.
- *Rudolph, S., & Gardner, M. K. (1986-1987). Remote site instruction in physics: A test of the effectiveness of a new teaching technology. *Journal of Educational Technology Systems*, 15(1), 61-80.
- Russell, T. L. (1999). *The no significant difference phenomenon*. Chapel Hill, NC: Office of Instructional Telecommunications, North Carolina State University.
- *Ryan, W. F. (1996). *The distance education delivery of senior high advanced mathematics courses in the province of Newfoundland and Labrador: A study of the academic progress of the participating students*. Unpublished doctoral dissertation, Ohio University, Athens, OH.
- *Ryan, W. F. (1996). The effectiveness of traditional vs. audiographics delivery in senior high advanced mathematics courses. *Journal of Distance Education*, 11(2), 45-55.
- Saba, F. (2000). Research in distance education: A status report. *International Review of Research in Open and Distance Education*, 1(1), 1-9.
- *Sankar, C. S., Ford, F. N., & Terase, N. (1998). Impact of videoconferencing in teaching an introductory MIS course. *Journal of Educational Technology Systems*, 26(1), 67-85.
- *Sankaran, S. R., & Bui, T. (2001). Impact of learning strategies and motivation on performance: A study in Web-based instruction. *Journal of Instructional Psychology*, 28(3), 191-198.
- *Sankaran, S. R., Sankaran, D., & Bui, T. X. (2000). Effect of student attitude to course format on learning performance: An empirical study in Web vs. lecture instruction. *Journal of Instructional Psychology*, 27(1), 66-73.
- Schlosser, C. A., & Anderson, M. L. (1994). *Distance education: Review of the literature*. Washington, DC: Association for Educational Communications and Technology.
- Schoenfeld-Tacher, R., & Persichitte, K. A. (2000). Differential skills and competencies required of faculty teaching distance education courses. *International Journal of Educational Technology*, 2(1), 1-16.
- *Schoenfeld-Tacher, R., & McConnell, S. (2001, April). *An examination of the outcomes of a distance-delivered science course*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.
- *Schoenfeld-Tacher, R., McConnell, S., & Graham, M. (2001). Do no harm — a comparison of the effects of on-line vs. traditional delivery media on a science course. *Journal of Science Education and Technology*, 10(3), 257-265.
- *Schulman, A. H., & Sims, R. L. (1999). Learning in an online format versus an in-class format: An experimental study. *THE Journal Online*, 26(11). Retrieved April 30, 2003 from

Bernard et al.

- <http://www.thejournal.com/magazine/>
- *Schutte, J. G. (1997). *Virtual teaching in higher education: The new intellectual superhighway or just another traffic jam?* Retrieved November, 2000 from <http://www.csun.edu/sociology/virexp.htm>
- *Scott, M. (1990). *A comparison of achievement between college students attending traditional and television course presentations (distance education)*. Unpublished doctoral dissertation, Auburn University, Auburn, AL.
- *Searcy, R. (1993). *Grade distribution study: Telecourses vs. traditional courses*. (Study prepared for the Calhoun Community College Steering Committee). (ERIC Document Reproduction Service No. ED 362 251)
- Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, 105, 309-316.
- Shachar, M. (2002). Differences between traditional and distance education outcomes: A meta-analytic approach. Unpublished dissertation, Touro University International.
- Shachar, M., & Neumann, Y. (2003). Differences between traditional and distance education academic performances: A meta-analytic approach. *International Review of Research in Open and Distance Education*, October. Retrieved October 30, 2003 from <http://www.irrodl.org/content/v4.2/shachar-neumann.html>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company.
- Shale, D. (1990). Toward a reconceptualization of distance education. In M. G. Moore (Ed.), *Contemporary issues in American distance education* (pp. 333-343). Oxford: Pergamon Press.
- *Simpson, H., Pugh, H. L., & Parchman, S. W. (1991). An experimental two-way video teletraining system: Design, development and evaluation. *Distance Education*, 12(2), 209-231.
- *Simpson, H., Pugh, H. L., & Parchman, S. W. (1993). Empirical comparison of alternative instructional TV technologies. *Distance Education*, 14(1), 147-164.
- *Sipusic, M. J., Pannoni, R. L., Smith, R. B., Dutra, J., Gibbons, J. F., & Sutherland, W. R. (1999). *Virtual collaborative learning: A comparison between face-to-face tutored video (TVI) and distributed tutored video instruction (DTV)*. Sun Microsystems, Inc. (In the SML technical report series 1999).
- *Sisung, N. J. (1992). *The effects of two modes of instructional delivery: Two-way forward facing interactive television and traditional classroom on attitudes, motivation, on-task/off-task behavior and final exam grades of students enrolled in humanities courses*. Unpublished doctoral dissertation, University of Michigan.
- *Smeaton, A. F., & Keogh, G. (1999). *An analysis of the use of virtual delivery of undergraduate lectures*. Retrieved April 30, 2003 from <http://citeseer.nj.nec.com/cache/papers/cs/5005/http://zSzzSzwww.compapp.dcu.ie/zSz~asmeatonzSzpubszSzCompEd98.pdf/an-analysis-of-the.pdf>
- *Smith, D. L., & McNelis, M. J. (1993, April). *Distance education: Graduate student attitudes and academic performance*. Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA.
- Smith, P. L., & Dillon, C. L. (1999). Comparing distance learning and classroom learning: Conceptual considerations. *American Journal of Distance Education*, 13, 107-124.
- *Smith, R. E. (1990). *Effectiveness of the interactive satellite method in the teaching of first year German: A look at selected high schools in Arkansas and Mississippi*. Unpublished

Bernard et al.

- doctoral dissertation, University of Mississippi, Oxford.
- *Smith, T. E. (2001). *A comparison of achievement between community college students attending traditional and video course presentations*. Unpublished doctoral dissertation, Auburn University, Auburn, AL.
- *Smith, T. L., Ruocco, A., & Jansen, B. J. (1999). *Digital video in education*. Proceeding of the ACM Computer Science Education Conference, 122-126, New Orleans, LA.
- *Sorensen, C. K. (1996). *Students near and far: Differences in perceptions of community college students taking interactive television classes at origination and remote sites*. (ERIC Document Reproduction Service No. ED 393 509)
- *Souder, W. E. (1993). The effectiveness of traditional vs. satellite delivery in three management of technology master's degree programs. *The American Journal of Distance Education*, 7(1), 37-53.
- Spector, J. M. (2001). Competencies for online teaching. *Eric Digest*. Report Number: EDO-IR-2001-09.
- *Spooner, F., Jordan, L., Algozzine, B., & Spooner, M. (1999). Student ratings of instruction in distance learning and on-campus classes. *Journal of Educational Research*, 92(3), 132-140.
- *Stone, H. R. (1990). *Does interactivity matter in video-based off-campus graduate engineering education?* Washington, DC: American Society for Engineering Education, Center for Professional Development. (ERIC Document Reproduction Service no. ED 317 421)
- *Summers, M., Anderson, J. L., Hines, A. R., Gelder, B. C., & Dean, R. S. (1996). The camera adds more than pounds: Gender differences in course satisfaction for campus and distance learning students. *Journal of Research and Development in Education*, 29(4), 212-229.
- *Suter, N. W., & Perry, M. K. (1997, November 12-14). *Evaluation by electronic mail*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Memphis, TN.
- Taylor, J. C. (2001). *Fifth generation distance education*. Keynote address delivered at the ICDE 20th World Conference, Dusseldorf, Germany, 1-5 April. Retrieved July 24, 2001 from <http://www.usq.edu.au/users/taylorj/conferences.htm>
- Tennyson, R. D. (1994). The big wrench vs. integrated approaches: The great media debate. *Educational Technology, Research and Development*, 42(2), 15-28.
- *Thirunarayanan, M. O., & Perez-Prado, A. (2001). Comparing Web-based and classroom-based learning: A quantitative study. *Journal of Research on Computing in Education*, 34(2), 131-137.
- *Thomerson, J. D. (1995). *Student perceptions of the affective experiences encountered in distance learning courses (interactive television)*. Unpublished doctoral dissertation, University of Georgia, Athens.
- *Tidewater Community College. (2001). *Distance learning report*. Norfolk, VA: Tidewater Community College, Office of Institutional Effectiveness.
- *Tiene, D. (1997). Student perspective on distance learning with interactive television. *TechTrends*, 42(1), 41-47.
- *Toussaint, D. (1990). *Fleximode: Within Western Australia TAFE*. Leabrook, Australia: TAFE National Centre for Research and Development, Ltd.(ERIC Document Reproduction Service No. ED 333 227)
- *Tucker, S. (2001). *Distance education: Better, worse or as good as traditional education?* Retrieved April 30, 2003 from www.westga.edu/~distance/ojdla/winter44/tucker44.html
- Ullmer, E. J. (1994). Media and learning: Are there two kinds of truth? *Educational Technology*

Bernard et al.

Research & Development, 42 (1), 21-32.

- *Umble, K. E., Cervero, R. M., Yang, B., & Atkinson, W. L. (2000). Effects of traditional classroom and distance continuing education: a theory-driven evaluation of a vaccine-preventable diseases course. *American Journal of Public Health*, 90(8), 1218-1224.
- Ungerleider, C., & Burns, T. (2003). *A systematic review of the effectiveness and efficiency of networked ICT in education: A state of the art report to the Council of Ministers Canada and Industry Canada*. Ottawa, ON: Industry Canada.
- Valentine, J. C., & Cooper, H. (2003). *What works clearinghouse study design and implementation assessment device* (Version 1.0). Washington, DC: U.S. Department of Education.
- Verduin, J. R., & Clark, T. A. (1991). *Distance education: The foundations of effective practice*. San Francisco, CA: Jossey-Bass.
- *Waldmann, E., & De Lange, P. (1996). Performance of business undergraduates studying through open learning: A comparative analysis. *Accounting Education*, 5(1), 25-33.
- *Walker, B. M., & Donaldson, J. F. (1989). Continuing engineering education by electronic blackboard and videotape: A comparison of on-campus and off-campus student performance. *IEEE Transactions on Education*, 32(4), 443-447.
- *Wallace, L. F., & Radjenovic, D. (1996). *Remote training for school teachers of children with diabetes mellitus*. Retrieved September 13, 2001 from <http://www.unb.ca/naweb/proceedings/1996/zwallace.html>
- *Wallace, P. E., & Clariana, R. B. (2000). Achievement predictors for a computer-applications module delivered online. *Journal of Information Systems Education*, 11(1/2), 13-18. Retrieved May 15, 2003 from <http://gise.org/JISE/Vol11/v11n1-2p13-18.pdf>
- *Wang, A. Y., & Newlin, M. H. (2000). Characteristics of students who enroll and succeed in psychology web-based classes. *Journal of Educational Psychology*, 92(1), 137-143.
- *Waschull, S. B. (2001). The online delivery of psychology courses: Attrition, performance, and evaluation. *Teaching of Psychology*, 28(2), 143-146.
- *Wegner, S. B., Holloway, K. C., & Garton, E. M. (1999, November). The effects of Internet-based instruction on student learning. *Journal of Asynchronous Learning Networks*, 3(2), Retrieved May 31, 2002, from http://www.aln.org/alnweb/journal/Vol3_issue2/Wegner.htm.
- Weiner, B. (1992). *Human motivation: Metaphors, theories and Research*. Newbury Park CA: Sage Publications.
- *Westbrook, T. S. (1998). Changes in students' attitude toward graduate business instruction via interactive television. *The American Journal of Distance Education*, 11(1), 55-69.
- *Whetzel, D. L., Felker, D. B., & Williams, K. M. (1996). A real world comparison of the effectiveness of satellite training and classroom training. *Educational Technology Research and Development*, 44(3), 5-18.
- *Whitten, P., Ford, D. J., Davis, N., Speicher, R., & Collins, B. (1998). Comparison of face-to-face versus interactive video continuing medical education delivery modalities. *Journal of Continuing Education in the Health Professions*, 18(2), 93-99.
- *Wick, W. R. (1997). *An analysis of the effectiveness of distance learning at remote sites versus on-site locations in high school foreign language programs*. Unpublished doctoral dissertation, University of Minnesota.
- *Wideman, H. H., & Owston, R. D. (1999). *Internet based courses at Atkinson College: An initial assessment*. Retrieved May 13, 2003 from <http://www.yorku.ca/irlt/reports/techreport99-1.htm>
- Winkelmann, C.L. (1995). Electronic literacy, critical Pedagogy, and collaboration: A case for

Bernard et al.

cyborg writing. *Computers and the Humanities*, 29(6), 431-448.

- *Winn, F. J., Fletcher, D., Smith, J., Williams, R., & Louis, T. (1999). Internet teaching of PA practitioners in rural areas: Can complex material with high definition graphics be taught using PC. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting*.
- *Wisher, R. A., & Priest, A. N. (1998). Cost-effectiveness of audio teletraining for the U.S. Army National Guard. *The American Journal of Distance Education*, 12(1), 38-51.
- *Wishner, R. A., Curnow, C. K., & Seidel, R. J. (2001). Knowledge retention as a latent outcome measure in distance learning. *The American Journal of Distance Education*, 15(3), 20-3.

Appendix

Coded variables and study features in the DE meta-analysis codebook

Section A: Identification of Studies

1. Study Number (Name: "Study"):
2. Finding Number (Name: "Finding"):
3. Author Name (Name: "Author"):
4. Year of Publication (Name: "Yr"):

Section B: Outcome Features

1. Outcome Type (Name: "Outcome"):
 1. Achievement
 2. Retention
 3. Attitude towards course
 4. Attitude towards the technology
 5. Attitude towards the subject matter
 6. Attitude towards the instructor
 7. Other attitudes
2. Whose Outcome (Name: "Whose"):
 1. Group
 2. Individual
 3. Teacher
3. Number of Control Conditions (Name: "Ctrl"):
 1. One control, one DE
 2. One control, more than one DE
 3. One DE, more than one control
 4. More than one DE and more than one control

Section C: Methodological Features

1. Type of Publication (Name: "Typpub"):
 1. Journal article
 2. Book chapter
 3. Report
 4. Dissertation
 2. Outcome Measure (Name: "Measure"):
 1. Standardized test
 2. Researcher-made test
 3. Teacher-made test
 4. Teacher/researcher-made test
 3. Effect Size (Name: "Esest"):
 1. Calculated
 2. Estimated from probability levels
 4. Treatment Duration (Name: "Durat"):
 1. Less than one semester
 2. One semester
 3. More than one semester
 5. Treatment Proximity (Name: "Prox"):
 1. Same time period
-

-
2. Different time period
 6. Instructor equivalence (Name: "Inseq"):
 1. Same instructor
 2. Different instructor
 7. Student equivalence (Name: "Stueq"):
 1. Random assignment
 2. Statistical control
 8. Equivalent time on task (Name: "Timeeq"):*
 9. Material Equivalence (Name: "Mateq"):
 1. Same curriculum materials
 2. Different curriculum materials
 10. Learner ability (Name: "Abilit"):*
 11. Attrition Rates (Name: "Attr"):*
 12. Average Class Size (Name: "Size"):
 1. DE larger than control
 2. DE equal to control
 3. DE smaller than control
 13. Gender (Name: "Gender"):*

Section D: Course Design and Pedagogical Features

1. Simultaneous Delivery (Name: "Simul"):
 1. Simultaneous delivery
 2. Not simultaneous
2. Systematic ID "Instructional Design" (Name: "Id"):*
3. DE condition: Advance Information (Name: "Advinf"):
 1. Information received prior to commencement of the course
 2. Information received at the first course
 3. No information received
4. Opportunity for F2F/instructor (Name: "f2ft"):
 1. Yes. Opportunity to meet the instructor *during* instruction
 2. No opportunity to meet the instructor
 3. Yes. Opportunity to meet the instructor prior to, or at the commencement of, instruction only (example: orientation session)
5. Opportunity for F2F contact/peers (Name: "f2fp"):
 1. Yes Opportunity to meet peers during instruction
 2. No opportunity to meet peers
 3. Opportunity to meet peers at or prior to the commencement of instruction
6. Provision for Synchronous technically-mediated Communication /teacher (Name: "Syncte"):
 1. Opportunity for synchronous communication
 2. No opportunity for synchronous communication
7. Provision for synchronous technically-mediated communication/students (Name: "Synper"):
 1. Opportunity for synchronous communication
 2. No opportunity for synchronous communication
8. Teacher/Student Contact Encouraged (Name: "Tstd"):*
9. Student/Student Contact Encouraged (Name: "Ss"):*
10. Problem-based Learning (Name: "Pbl"):*

Section E: Institutional Features

1. Institutional Support for Instructor (Name: "Insupp"):*
-

2. Technical Support for Students (Name: "Tcsupp"):*

Section F: Media Features

1. Use of 2-way audio conferencing (Name: "Ac"):*
2. Use of 2-way video conferencing (Name: "Vc"):*
3. Use of CMC or interactive computer classroom (Name: "Cmc"):*
4. Use of e-mail (Name: "E-mail"):*
5. Use of 1-way broadcast TV or videotape or audiotape (Name: "Tvvid"):*
6. Use of web-based course materials (Name: "Web"):*
7. Use of telephone (Name: "Tele"):*
8. Use of computer-based tutorials/simulations (Name: "Cbi"):*

Section G: Demographics

1. Cost of Course Delivery (Name: "Cost"): *
 2. Purpose for Offering DE (Name: "Purpos"):
 1. Flexibility of schedule or travel
 2. Preferred media approach
 3. Access to expertise (teacher/program)
 4. Special needs students
 5. Efficient delivery or cost savings
 6. Multiple reasons. Specify:
 3. Instructor Experience with DE (Name: "Inde"):
 1. Yes
 2. No
 4. Instructor Experience with technologies used (Name: "Intech"):
 1. Yes
 2. No
 5. Students' Experience with DE (Name: "Stude"):
 1. Yes
 2. No
 6. Students' Experience with technologies used (Name: "Stutech"):
 1. Yes
 2. No
 7. Types of Control Learners (Name: "Lrtpc"):
 1. Grade School (K-12)
 2. Undergraduate
 3. Graduate
 4. Military
 5. Industry/business
 6. Professionals (e.g., doctors)
 8. Types of DE Learners (Name: "Lrtpd"):
 1. Grade and high school (K-12)
 2. Undergraduate
 3. Graduate
 4. Military
 5. Industry/business
 6. Professionals (e.g., doctors)
 9. Setting (Name: "Settng"):
 1. DE urban and control rural
-

-
2. DE urban and control urban
 3. DE reported/control not reported
 4. DE rural and control urban
 5. DE rural and control rural
 6. Control reported/DE not reported

10. Subject Matter (Name: "Subjec"):

1. Math (including stats and algebra)
2. Languages (includes language arts and second languages)
3. Science (including biology, sociology, psychology & philosophy)
4. History
5. Geography
6. Computer science (information technology)
7. Computer applications
8. Education
9. Medicine or nursing (histology)
10. Military training
11. Business
12. Engineering
13. Other (specify)

11. Average age (Name: "Age"):

1. Real difference in age means with the corresponding sign

* These items were coded using the following scheme:

1. DE more than control group
 2. DE reported/control group not reported
 3. DE equal to control group
 4. Control reported/DE not reported
 5. DE less than control group
- 999 Missing (no information on DE or control reported)

Bernard et al.

Author Notes

Robert M. Bernard is a Professor of Education at Concordia University in Montreal, Quebec, a member of the Centre for the Study of Learning and Performance (CSLP) and a Co-Chair of the Campbell Collaboration's Education Coordinating Group. He specializes in distance education, electronic publishing, research design and statistics, and quantitative synthesis.

Philip C. Abrami is a Professor, Research Chair, and Director of the Centre for the Study of Learning and Performance, Concordia University, Montreal, Quebec, Canada. He is currently a Co-Chair of the Education Coordinating Group of the Campbell Collaboration. His research interests include quantitative synthesis, technology integration in schools, and the social psychology of education.

Yiping Lou is an Assistant Professor of Educational Technology in the Department of Educational Leadership, Research, and Counseling at Louisiana State University, Baton Rouge, LA. Her current research interests include online collaborative learning, distance education, and models of technology integration.

Evgueni Borokhovski is a Doctoral Candidate in the Department of Psychology at Concordia University and a Research Assistant at the CSLP.

Anne Wade (M.L.I.S.) is Manager and Information Specialist at the CSLP. Her expertise is in information storage and retrieval, and research strategies. She is a lecturer in the Information Studies Program in the Department of Education, Concordia University; a member of Campbell's Information Retrieval Methods Group; and an Associate of the Evidence Network, UK.

Lori Wozney is a Doctoral Candidate in Concordia University's Educational Technology program and is a member of the CSLP. Her work focuses on the integration of instructional technology, self-regulated learning, and organizational analysis.

Peter Andrew Wallet is currently an M.A. student in Educational Technology at Concordia University. He also holds a Masters degree in Educational Psychology from McGill University and is currently working for the UNESCO Institute for Statistics on global, teacher-related statistics. Teacher training using distance education is a prime interest of his research.

Manon Fiset is a graduate of Concordia University's M.A in Educational Technology. She has worked in the field of distance education for several years at the Institute of Canadian Bankers, the CSLP, and is currently working as a Project Manager at Cegep@distance in Montreal.

Binru Huang is a research assistant in the Department of Educational Leadership, Research, and Counseling at Louisiana State University, Baton Rouge, LA. She is also a graduate student in the Department of Experimental Statistics. Her research interests include statistical analysis methods and the use of technology in distance learning.

Footnote

¹ We explored another method of entering the three sets of study features in blocks. First, we ran each block separately and saved the unstandardized predicted values (Y'). This provided three new composite variables, which were then entered into WMR in different orders, as indicated above. The results were very similar to the ones reported, with the exception that a clearer distinction emerged between pedagogy and media (i.e., pedagogy was always significant and media was never significant). However, we chose to report the results in the manner described above because it allows a detailed analysis of the contribution of individual study features, whereas the method just described does not. Also, neither synchronous nor asynchronous DE formed a homogeneous set.

