

Review of Educational Research

<http://rer.aera.net>

How Does Distance Education Compare With Classroom Instruction? A Meta-Analysis of the Empirical Literature

Robert M. Bernard, Philip C. Abrami, Yiping Lou, Evgueni Borokhovski, Anne Wade,
Lori Wozney, Peter Andrew Wallet, Manon Fiset and Binru Huang
REVIEW OF EDUCATIONAL RESEARCH 2004; 74; 379
DOI: 10.3102/00346543074003379

The online version of this article can be found at:
<http://rer.sagepub.com/cgi/content/abstract/74/3/379>

Published on behalf of



<http://www.aera.net>

By



<http://www.sagepublications.com>

Additional services and information for *Review of Educational Research* can be found at:

Email Alerts: <http://rer.aera.net/cgi/alerts>

Subscriptions: <http://rer.aera.net/subscriptions>

Reprints: <http://www.aera.net/reprints>

Permissions: <http://www.aera.net/permissions>

How Does Distance Education Compare With Classroom Instruction? A Meta-Analysis of the Empirical Literature

Robert M. Bernard and Philip C. Abrami

Concordia University

Yiping Lou

Louisiana State University

Evgueni Borokhovski, Anne Wade, Lori Wozney,

Peter Andrew Wallet, and Manon Fiset

Concordia University

Binru Huang

Louisiana State University

A meta-analysis of the comparative distance education (DE) literature between 1985 and 2002 was conducted. In total, 232 studies containing 688 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 that many 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 effect sizes for asynchronous applications favored DE. However, significant heterogeneity remained in each subset.

KEYWORDS: classroom instruction, comparative studies, distance education, meta-analysis, research methodology.

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 1960s, 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 1980s), multimedia (1980s and 1990s), teleconferencing (1990s), and distance education (DE) (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.

The present study dealt specifically with comparative studies of DE. Keegan's (1996) definition of 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 in planning, preparation, and 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 on teleconferencing technologies that are group based. Some argue that when DE simply re-creates 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, p. 1) 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)

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, delivery of materials, 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 multimedia, the Internet, access to Web-based resources, computer-mediated communication (CMC), 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. This literature was the focus of the present 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 (Kozma, 1994; Morrison, 1994; Tennyson, 1994; Ullmer, 1994). Kozma argued that Clark’s original assessment was based on “old non-interactive technologies” that simply carried method and content, wherein 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 nonhuman agents or tools, so 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; rather, it depends entirely on the availability of media for delivery and communication (Keegan, 1996). DE can be noninteractive or highly interactive and may, in fact, encompass one or many media types (e.g., print, video, computer-based simulations, and 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 executed comparative study in which 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 that 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 in terms of solving complex problems may develop through peer modeling and mentoring (Lou, 2004; Lou, Dedic, & 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 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 that would render such studies uninterpretable.

There may be a more practical reason for assessing the effectiveness of DE in comparison with 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, television coverage), DE itself tended to be restricted by geographical boundaries (e.g., for many years the

United Kingdom 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 Internet service providers, has set traditional educational institutions into intense competition for the worldwide market of “online learners.” So it is arguable that finding answers to the question that has guided much of the comparative 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? According to long-standing 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 simply representing “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 a 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 the quality of the research itself and the veracity of its findings can be assessed.

There have been many attempts to summarize the comparative DE research literature. The most comprehensive, but least assiduous, is Russell’s (1999) collection of 355 “no significant difference” studies. On the basis of compiling 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 evidence with the much smaller number of “significant difference studies” (which could 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 that, by extension, 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 means only that they do not exist in the sample being studied. This is particularly true in small-sample studies wherein the power to reject the null hypothesis (and thus the risk of making Type II errors) is high. Third, the different sample sizes of individual studies make it impossible to aggregate the results of different studies solely on the basis of their test statistics. Thus, 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 titled *What's the Difference? A Review of Contemporary Research on the Effectiveness of Distance Learning in Higher Education*, may contain a level of bias similar 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 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 may generally be true—that the literature lacks rigor of methodology and reporting—the finding of the "questionable effectiveness of DE" based on a select number of studies is no more credible than Russell's claim of nonsignificance based on everything that has ever been published. Somewhere between these extremes resides 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; Jung & Rha, 2000; Moore & Thompson, 1990; Saba, 2000; Schlosser & Anderson, 1994). This type of research has long been known for its subjectivity, potential bias, and inability to answer questions about magnitudes 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 that can be averaged in a way that test statistics cannot. Refinements made by Hedges and Olkins (1985) further reduced 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, along with the associated variability. 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 on DE studies can allow us to estimate the 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, and so forth. Research methodology can also be investigated, thereby 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 in terms of both quantity and quality, indicating directions for further inquiry.

In summary, meta-analysis has the following advantages: (a) It answers questions about sizes of effects; (b) it allows systematic exploration of sources of variability in effect sizes; (c) it allows for control over *internal validity* by focusing on

comparison studies versus 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 are analyzed; (f) it uses the student as the unit of analysis, not the study (large sample studies have higher weights); (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 reanalysis of parts of the data set for special purposes (e.g., military studies, synchronous versus asynchronous instruction, Web-based instruction); and (j) it allows comment on what we know and what we need to know (Abrami, Cohen, & d'Apollonia, 1988; Bernard & Naidu, 1990).

Five quantitative syntheses specifically related to DE and its correlates have been published (Allen, Bourhis, Burrell, & Mabry, 2002; Cavanaugh, 2001; Machtmes & Asher, 2000; Shachar & Neumann, 2003; Ungerleider & Burns, 2003). 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 offered by other studies that DE and classroom instruction are relatively equal. In another recent study, Ungerleider and Burns conducted a systematic review for the Council of Ministers of Education of Canada including a quantitative meta-analysis of the literature on networked and online learning (i.e., not specifically DE). They found poor methodological quality, to the extent that only 12 achievement and 4 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. This provides an example of two credible works offering conflicting evidence as to the state of comparative studies.

Allen et al. (2002) 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 ($r = .031$, $k = 25$, $N = 4,702$; significantly heterogeneous sample) favoring classroom instruction. When three outliers were removed from the analysis, the correlation coefficient increased to $.090$, and the homogeneity assumption was satisfied. Virtually no effects were found for "channel of communication" (video, audio, or written) or its interaction with "availability of interaction." This meta-analysis was limited in that it investigated 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 yielded little more than basic information related to the question of DE effectiveness.

The Cavanaugh (2001) 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 was 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 (2000), compared live or preproduced adult telecourses with their classroom equivalents on measures of classroom achievement in either experimental or quasi-experimental designs. Of 30 studies identified, 19 dated between 1943 and 1997 were coded for effect sizes and study features. The overall weighted effect size for these comparisons was -0.0093 (not significant; range: -0.005 to 1.50). The assumption of homogeneity of effect size was violated, and this 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; we also find great variability among the 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 computer-mediated communication (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 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., Pennsylvania State University) that began in the late 1940s. The primary purpose of this movement in the United States 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 and 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, in which students in remote locations work independently or in asynchronous groups, usually with the support of an instructor or tutor. We call 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, wherein learners were truly independent, connected to an instructor or tutor by the postal system; communication was truly asynchronous because of postal delays. Because of the differences in synchronous and asynchronous DE just noted, we decided to examine these two patterns undivided as well as 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 is only 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 in which both synchronous and asynchronous forms of communication are available to the learner. We have included both types in our review to examine how synchronicity and asynchronicity affect learning. When a choice in instructional design exists, knowing the influence of these patterns may guide the design. When there is no choice in design and students must learn asynchronously, separated in both space and time, it may be necessary to develop new instructional resources as alternative supports for student learning needs.

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. Thus, within asynchronous studies there is an element of within-group synchronicity (i.e., DE students communicating, synchronously, among themselves), just as there is a certain degree of 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 was to provide an exhaustive quantitative synthesis of the comparative research literature on DE, from 1985 to the end of 2002, across all age groups, media types, instructional methods, and outcome measures. From this literature, we sought 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 with 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?
7. What are important implications for practice and future directions for research?

Method

This meta-analysis was a quantitative synthesis of empirical studies conducted since 1985 that compared the effects of DE and traditional classroom-based instruction on student achievement, attitude, and retention (i.e., opposite of dropout). The year 1985 was chosen as a cutoff date because electronically mediated, interactive DE became widely available around that time. The procedures employed in conducting this quantitative synthesis are described subsequently 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 feature coding, and data analysis. (See Appendix A for a description of the variables and study features used in the final coding.)

Working Definition of DE

Our 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 Web resources). Generations 3, 4, and 5 represent moves away from directed and noninteractive 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 is thus as follows:

- Semipermanent separation (place and/or time) of learner and instructor during planned learning events.
- Presence of planning and preparation of learning materials, student support services, and final recognition of course completion by an educational organization.
- 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 criteria:

How Does Distance Education Compare With Classroom Instruction?

1. It had to 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. It had to 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 in which electronic media were used to supplement regular face-to-face classes with the teacher physically present were excluded.
3. It had to 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, inferential statistics, or sample size) were excluded.
4. It had to be publicly available or archived.
5. It had to include at least one achievement, attitude, or retention outcome measure.
6. It had to include an identifiable level of learner. All levels of learners from kindergarteners to adults, whether involved in informal schooling or professional training, were admissible.
7. It had to be published or presented no earlier than 1985 and no later than December of 2002.
8. It had to include outcome measures that were the same or comparable. If the study explicitly indicated that different exams were used for the experimental and control groups, the study was excluded.
9. It had to include outcome measures that reflected individual courses rather than entire programs. Thus, programs composed of many different courses, in which no opportunity existed to analyze conditions and corresponding outcomes for individual treatments, were excluded.
10. It had to 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 2002. Electronic searches were performed via 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 searches were performed with the Google, AlltheWeb, and Teoma search engines. Manual searches were performed in ComAbstracts, Educational Technology Abstracts; in several distance learning journals, including the *American Journal of Distance Education*, *Distance Education*, the *Journal of Distance Education*, *Open Learning*, and the *Journal of Telemedicine and Telecare*; and in several conference proceedings, including the Association for the Advancement of Computing in

Education, the American Educational Research Association, the Canadian Association for Distance Education, EdMedia, E-Learn, SITE, and WebNet. In addition, the reference lists of several earlier reviews, including those of Moore and Thompson (1990), Russell (1999), Machtmes and Asher (2000), Cavanaugh (2001), Allen et al. (2002), and Shachar (2002), were searched 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 items retrieved. Each of the studies retrieved was read by two researchers for possible inclusion according to the inclusion/exclusion criteria. The initial interrater agreement as to inclusion was 89%. Any study that was considered for exclusion by one researcher was cross-checked by another researcher. Two hundred thirty-two studies met all inclusion criteria and were included in this meta-analysis; 630 were excluded. The categories of reasons for exclusion and the numbers and percentages of excluded studies are shown in Appendix B.

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 evaluations of the course as a whole, the instructor, the course 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:

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

Cohen’s d values were converted to Hedges’ g values (i.e., unbiased estimates) via Equation 2 (Hedges & Olkin, 1985, 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 et al. (1981) and Hedges, Shymansky, and Woodworth (1989). These effect sizes were referred to in coding as “estimated effect sizes.” The following rules governed calculation of effect sizes:

- When multiple achievement data were reported (e.g., assignments, midterm and final exams, grade point averages, grade distributions), final exam scores were used in calculating effect sizes.
- When there was more than one control group and groups 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 elements of DE treatment (e.g., originating studio site), the former was used as the control group.
- In studies in which there were two DE conditions and one control condition, the weighted average of the two DE conditions was used.
- In studies in which 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 toward subject matter) so that only one effect size was generated from each study for each outcome.
- In the case of 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.
- When the direction was reported, a “midpoint” approach was taken to estimate a representative *t* value (i.e., midpoint between zero 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. Inter-coder agreement rates were 91% for number of effect sizes extracted within studies and 96% for effect size calculations. In total, 688 independent effect sizes (i.e., 321 achievement outcomes, 262 attitude outcomes, and 105 retention outcomes) were extracted.

Study Feature Coding

Initial coding. A comprehensive codebook was initially developed on the basis of several earlier narrative reviews (e.g., Phipps & Merisotis, 1999), meta-analyses (e.g., Cavanaugh, 2001), conceptual reports (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 dropped (e.g., type of student learning) if there were insufficient data in the primary literature for inclusion in the meta-analysis. As mentioned earlier, the variables and study features used in the final coding are described in Appendix A. In addition to these codes, elaborate operational descriptions were developed for each item and used to guide coders.

Operational definitions of coding options. 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, allowing us to quantify certain aspects of study features (i.e., methodology, pedagogy, and media) that have heretofore been ignored or dealt with qualitatively. Thus, we hoped that the meta-analysis would allow us to address the long-standing 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 entire research team adjudicated some difficult cases.

Synchronous and asynchronous DE. Outcomes were split, for the purposes 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 there was no synchronization with a classroom. As a result of this definition, some DE students did communicate synchronously with instructors or other students, but this was not typically the case. We did not separate conditions in which inter-DE synchronous communication occurred from those in which it did not. Outcomes for which “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 1. 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: (a) 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); (b) recode missing data using a mean substitution procedure under the assumption that missing data were “typical” of the average for each study feature; or (c) code missing data as zero under the assumption that these data also represented indetermination. We chose the last of these three options. The coded study features were (a) type of publication, (b) type of measure, (c) effect size (i.e., calculated or estimated), (d) treatment duration, (e) treatment time proximity, (f) instructor equivalence, (g) selection bias, (h) time-on-task equivalence, (i) material equivalence, (j) learner ability equivalence, (k) mortality, (l) class size equivalence, and (m) 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 similar to that used to produce the methodological study features, we recoded pedagogical and media-related study features 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 it 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 described subsequently 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 through the use of the weighting system, also shown in Table 1.

The nine pedagogical coded study features were as follows: (a) systematic instructional design procedures used, (b) advance course information given to students, (c) opportunity for face-to-face (F2F) contact with the teacher, (d) opportunity for F2F contact among student peers, (e) opportunity for mediated communication (e.g., e-mail, CMC) with the teacher, (f) opportunity for mediated communication among students, (g) student/teacher contact encouraged through activities or course design, (h) student/student contact encouraged through activities or course design, and (i) use of problem-based learning. The media-related items were as follows: (a) use of two-way audio conferencing, (b) use of two-way video-

TABLE 1
*Methodological, pedagogical, and media study feature codes
 and the recodes assigned to them*

Study feature code	Methodology recode	Pedagogy/ media recode
1. DE more than control group	-2	+2
2. DE reported/control group not reported	0	+1
3. DE equal to control group	+2	0
4. Control reported/DE not reported	0	-1
5. DE less than control group	-2	-2
999. Missing (no information or DE control reported)	0	0

conferencing, (c) use of CMC, (d) use of e-mail, (e) use of one-way TV or video- or audiotape, (f) use of the Web, (g) use of a telephone, and (h) use of computer-based instruction.

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.

In the case of one study with retention outcomes (Hittelman, 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 sample size reduced proportionally. The treatment k was then proportionally weighted. This procedure was used to avoid overweighting by one study. Outlier analyses were performed with the homogeneity statistic 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. The statistic used, Q_w , represents an extremely sensitive test of the homogeneity assumption and is evaluated via the chi-square sampling distribution.

To determine whether the findings for each mean outcome shared a common effect size, we tested the set of effect sizes for homogeneity with the homogeneity statistic Q_T . When all findings share the same population effect size, Q_T has an approximate chi-square 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 chi-square distribution with $p - 1$ degrees of freedom, where p is the number of classes. If Q_B is greater than the critical value, this 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 chi-square 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, this indicates that the effect sizes within the class are heterogeneous. We conducted data analyses using Comprehensive Meta-Analysis (Biostat) and SPSS (Version 11 for the Macintosh OS X).

Multiple regression modeling of study features. Weighted multiple regression in SPSS was used to explore variability in effect sizes and to model the relationships that existed 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 variance, and Equation 4 was used in calculating the weighting factor (Hedges & Olkin, 1985, p. 174).

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

and

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

Each set of study features, methodological, pedagogical, and media, was entered into weighted multiple regression analyses separately in blocks with 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. The individual beta value for each predictor was used in testing the significance of individual study features, and standard errors were corrected according to Equation 5 (Hedges & Olkin, 1985, p. 174).

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

Ninety-five percent confidence intervals (CIs) were corrected according to Equation 6 (Hedges & Olkin, 1985, p. 171).

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

We created the test statistic z to test the null hypothesis that $\beta = 0$ (via Equation 7), and we evaluated α using $t = 1.96$ (Hedges & Olkin, 1985, p. 172).

$$z_{\beta} = \frac{\beta}{\sqrt{\sigma_{\beta}}} \tag{7}$$

Results

In total, 232 studies yielding 688 independent effect sizes (i.e., outcomes) were analyzed. These values were based on totals of 57,019 students ($k = 321$) with achievement outcomes, 35,365 students ($k = 262$) with attitude outcomes, and 57,916,029 students ($k = 105$) with retention outcomes. The sample size 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 in calculating 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 2 shows a breakdown of missing study feature data over the three outcome measures: achievement, retention, and attitude. 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 does not work in DE.

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 value was extreme (i.e., above 500). This left 318 achievement outcomes ($N = 54,775$) to be analyzed.

The tables in Appendix C show frequencies and percentages of achievement outcomes according to date of publication and source of publication. Two things

TABLE 2
Numbers and percentages of missing values for the three measures

Measure	Total cells	No. 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	55.74

are evident in these tables. First, the impetus to conduct comparative research 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 was $-.035$ ($df = 316, p < .05$), indicating that there was no systematic relationship between these two variables. Second, there is modest bias over the three classes of publication sources upon which these data were based. The $g+$ value for technical reports, while not substantially greater than that for dissertations, was significant.

Table 3 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.

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 , a skewness value of 0.203, and a kurtosis value of 0.752; the distribution is nearly normal. It is clear from the range of effect sizes (-1.31 to $+1.41$) that some applications of DE are far better than classroom instruction and that some are far worse.

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 ($g+$ values), standard errors, confidence intervals, and homogeneity statistics for these three categories are shown in Table 3. The difference in $g+$ resulting from this split, with synchronous DE significantly negative and asynchronous DE significantly positive, is dramatic, but both groups remained heterogeneous. Further exploration of variability in g is required.

Weighted multiple regression. In beginning to explore the variability in g , we conducted weighted multiple regression (WMR) analyses with the three blocks of predictors. We were particularly interested in the variance accounted for by each of

TABLE 3
Weighted mean effect sizes for combined achievement outcomes

	Effect size		95% confidence interval		Homogeneity of effect size	
	$g+$	SE	Lower	Upper	Q value	df
Combined outcomes ($k = 318, N = 54,775$)	0.0128	0.0100	-0.0068	0.0325	1,191.32*	317
Synchronous ($k = 92, N = 8,677$)	-0.1022*	0.0236	-0.1485	-0.0559	182.11*	91
Asynchronous ($k = 174, N = 36,531$)	0.0527*	0.0121	0.0289	0.0764	779.38*	173
Unclassified ($k = 52, N = 9,567$)	-0.0359	0.0273	-0.0895	0.0177	191.93*	51

* $p < .05$.

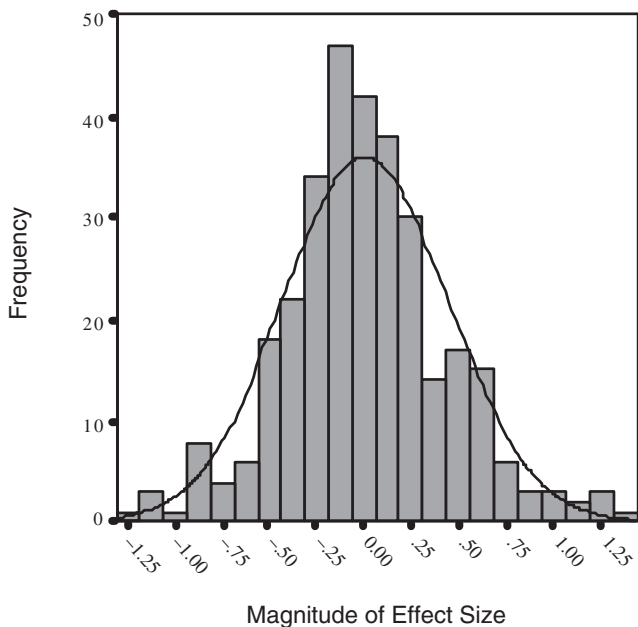


FIGURE 1. *Distribution of 318 achievement effect sizes*

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 effects attributable to features of pedagogy and media and that pedagogy and media themselves are confounded in studies of this type. In this analysis, we 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, was 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 accounted for the more general findings. We entered the three blocks of predictors¹ (e.g., 13 methodological study features) into the WMR in different orders: (a) methodology followed by pedagogy and media, (b) methodology followed by media and pedagogy, (c) pedagogy followed by media and methodology, and (d) media followed by pedagogy and methodology. We did not enter methodology in the second step because this combination seemed to explain little of interest. The partitioning of between-group (Q_B) and within-group (Q_W) variance in the third step of the regression for both synchronous and asynchronous DE outcomes yielded the following results: Q_B was significant for both DE patterns, and synchronous DE outcomes were homogeneous (i.e., Q_W was not significant) while asynchronous DE outcomes were not (i.e., Q_W was significant).

TABLE 4

Comparison of R² changes for blocks of study features: Achievement outcomes

Predictor	1st step	2nd step after methodology	2nd step after pedagogy or media	3rd step
Synchronous DE				
Methodology	.490*			.250*
Pedagogy	.360*	.101*	.130*	.077
Media	.245*	.058	.015	.048
Asynchronous DE				
Methodology	.117*			.054
Pedagogy	.156*	.107*	.124*	.120*
Media	.111*	.051	.078	.065

Note. Not all significance tests were based on the same degrees of freedom.

* $p < .05$.

Table 4 provides a comparison of 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., the third step in both cases), methodology and pedagogy were always significant, no matter which position they were in or whether outcomes were associated with synchronous or asynchronous DE. Second, media was significant only when it was entered in the first step. Overall, this indicates that methodology and pedagogy are more important than media in predicting achievement. Third, in line with much of the commentary on the research literature of DE and other media comparison literatures, research methodology accounted 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 according to fuzzy criteria such as “more than one methodological flaw.” By including studies ranging in methodological quality and coding for such differences, we overcame this difficulty to an extent.

Study feature analysis. We examined individual study features for pedagogy and media after variation for methodology had been accounted for, in order to determine which features had the greatest effect on achievement outcomes. The results of this analysis (i.e., the significant study features resulting from the WMR) are summarized in Table 5.

Demographic study features. We also coded a set of study features relating to demographics of students, instructors, subject matter, and reasons for offering DE. Appendix D contains the three study features that yielded enough outcomes to warrant analysis. DE achievement effects were large (a) when 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 ($g+ = 0.1777$). Interestingly, there was no difference between undergraduate postsecondary education applications of DE and classroom instruction. Graduate school applications yielded modest but significant results in favor of DE

TABLE 5

Summary of study features significantly predicting achievement, attitude, and retention outcomes

Favor classroom instruction (-)	Favor DE (+)
Synchronous DE	
<i>Achievement</i>	<i>Achievement</i>
<ul style="list-style-type: none"> • Face-to-face meetings with instructor • Use of 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>	<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 instructional design • Opportunity for mediated communication with instructor • Instructor/student contact encouraged • Use of telephone to contact instructor
<i>Retention</i>	<i>Retention</i>
<ul style="list-style-type: none"> • No significant predictors 	<ul style="list-style-type: none"> • No significant predictors
Asynchronous DE	
<i>Achievement</i>	<i>Achievement</i>
<ul style="list-style-type: none"> • No significant predictors 	<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
<i>Attitudes</i>	<i>Attitudes</i>
<ul style="list-style-type: none"> • Use of the Web 	<ul style="list-style-type: none"> • Use of problem-based learning strategies • Use of computer-mediated communication • Use of computer-based instruction
<i>Retention</i>	<i>Retention</i>
<ul style="list-style-type: none"> • No significant predictors 	<ul style="list-style-type: none"> • No significant predictors

($g+ = 0.0809$). As well, the academic subject areas of math, science, and engineering appeared to be best suited to the classroom ($g+ = -0.1026$), while subjects related to computing and military/business ($g+ > 0.17$) seemed to work well in distance education settings.

Attitude Outcomes

Synchronous and asynchronous outcomes. We found various forms of attitude measures in the literature that could be classified into four categories: attitude toward technology, attitude toward subject matter, attitude toward instructor, and attitude toward course. We also had 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 analyses in which 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 the analysis, one extremely high outlier was removed. This left 154 outcomes to be analyzed.

TABLE 6

Weighted mean effect sizes for combined, synchronous, and asynchronous attitude outcomes

DE category	Effect size		95% confidence interval		Homogeneity of effect size	
	<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 (<i>k</i> = 83, <i>N</i> = 9,483)	-0.1846*	0.0222	-0.2282	-0.1410	410.02*	82
Asynchronous (<i>k</i> = 71, <i>N</i> = 11,624)	-0.0034	0.0193	-0.0412	0.0344	345.64*	70

**p* < .001.

We split the sample into synchronous and asynchronous DE, in the same manner as for achievement, and found essentially the same overall dichotomy. Table 6 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 were negative, note the contrast between synchronous and asynchronous outcomes. The average effect size for synchronous outcomes was significant, while the average effect size for asynchronous outcomes was not. Furthermore, there was a high level of variability among effect sizes, even after the split. Figure 2 provides a graphic depiction of overall variabilities in attitude outcomes for 154 outcomes, and it can be seen that they ranged from -1.51 to +1.63. There were circumstances in which DE students' reactions were extremely positive, and others in which their reactions were quite negative, relative to classroom instruction.

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 indicated heterogeneity for these groups.

We examined R^2 changes for attitudes in regard to three blocks of predictors, methodology, pedagogy, and media, in different orders, in the same way we did for achievement outcomes. Table 7 presents a comparison of R^2 changes for blocks of study features entered in different orders in the 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 were more differences here between synchronous and asynchronous DE in the three blocks of predictors. As with achievement, methodology still accounted for more variation in synchronous DE than in asynchronous DE. While pedagogy was somewhat suppressed in the case of synchronous DE, it emerged as important in the case of asynchronous DE. On the other hand, media appeared to be more important in synchronous DE than in asynchronous DE.

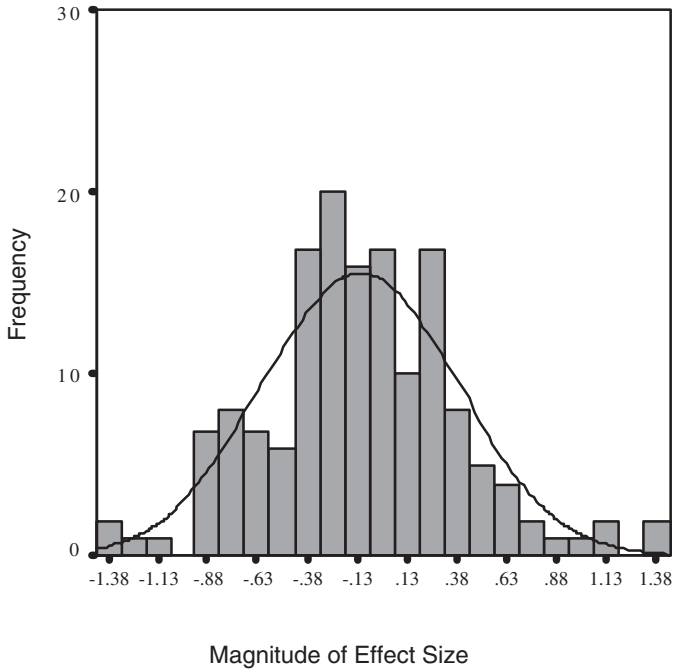


FIGURE 2. *Distribution of 154 attitude effect sizes*

TABLE 7
Comparison of R² changes for blocks of study features: Attitude outcomes

Predictor	1st step	2nd step after methodology	2nd step after pedagogy or media	3rd step
Synchronous DE				
Methodology	.471**			.421**
Pedagogy	.128	.138**	.101	.120**
Media	.136**	.067*	.109**	.049
Asynchronous DE				
Methodology	.218**			.157
Pedagogy	.253**	.215**	.133	.076
Media	.241**	.236**	.121	.097

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

Study feature analysis. Individual study features were assessed after the WMR in a manner similar to that for achievement outcomes; significant synchronous and asynchronous results are summarized in Table 5.

Retention Outcomes

Retention is defined here as the opposite of dropout or attrition. We found several statewide studies (e.g., California) comparing DE and classroom conditions in which 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 effect, 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 was wide variability.

Table 8 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 thus, 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 seemed to reflect the average for all studies. Caution should be exercised in interpreting the mean effect size for synchronous DE because of the low number of outcomes associated with it.

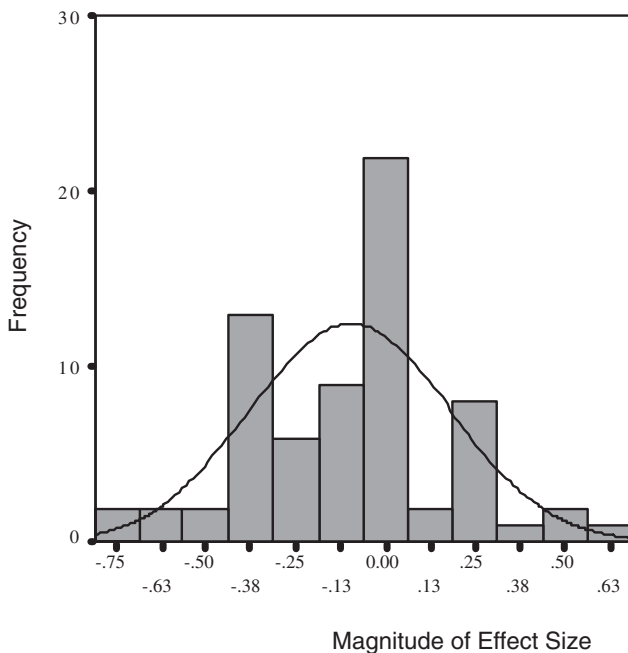


FIGURE 3. *Distribution of 70 retention effect sizes*

TABLE 8
Mean effect sizes for synchronous and asynchronous retention outcomes

Outcome type	Effect size		95% confidence interval		Homogeneity of effect size	
	<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 and poor student-teacher communication, we wondered whether this situation had changed over the years examined here as a result of the increasing availability of newer forms of electronic communication. To explore this issue, we calculated the Pearson product-moment correlation between dropout (i.e., *g*) and “year of publication” over the 17 years of the study. This correlation was .015 (*df* = 68, *p* > .05), suggesting that there was no systematic increase or decrease in differential retention rates over time. This situation was somewhat different for synchronous (*r* = -.27, *df* = 14, *p* > .05) and asynchronous (*r* = .011, *df* = 51, *p* > .05) retention outcomes calculated separately, although neither reached significance. Had the synchronous correlation been significant, this would have indicated a decreasing differential (i.e., the two conditions becoming more similar) over time between classroom and DE applications in terms of retention.

When a WMR analysis was performed on synchronous and asynchronous retention outcomes, the results for methodology, pedagogy, and media were all nonsignificant. Therefore, no regression outcomes are presented.

Summary of Results: Achievement

1. There was a very small and significant effect favoring DE conditions (*g*+ = 0.0128) in terms of overall achievement outcomes (*k* = 318). However, the variability surrounding this mean was wide and significant.
2. When outcomes were split between synchronous and asynchronous DE achievement outcomes, a small, significant negative effect (*g*+ = -0.1022) was found for synchronous DE, and a significantly positive effect was found 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 changes were examined for blocks of predictors entered in different orders, methodology and pedagogy were almost always found to be significant, whereas media was significant only when it was entered in the

first step. This was true for both synchronous DE outcomes and asynchronous DE outcomes. Individual significant study feature outcomes are summarized in Table 5.

Summary of Results: Attitude

1. There was a small negative but significant effect on overall attitude outcomes in favor of classroom instruction ($g+ = -0.0812$). Again, the variability around this mean was significantly heterogeneous.
2. There were differences in the effect sizes for synchronous DE ($g+ = -0.1846$) and asynchronous DE ($g+ = -0.0034$). Both favored classroom instruction, but the average effect size was significant for synchronous DE, and it was not for asynchronous DE. Individual significant study feature outcomes are summarized in Table 5.
3. R^2 change analyses of the type described earlier revealed that methodology, pedagogy, and media accounted for varying patterns of variance in terms of attitudes. It appears that these three sets of variables are related in a more complex way than they are for achievement outcomes.

Summary of Results: Retention

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

Summary of Results: Overall

1. There was extremely wide variability in effect size on all measures, and we were unable to find study features that formed homogeneous subsets, including the distinction between synchronous and asynchronous DE (with the one exception of synchronous DE in the case of achievement). This suggests that DE works extremely well sometimes and extremely poorly other times, even when all coded study features are taken into account.
2. Since the variation in effect size accounted for by methodology was 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 offering clear recommendations to practitioners and policymakers.
3. Another measure of the quality of the literature, amount of data available, suggested that the literature was very weak in terms of design features that would improve the interpretability of the results. More than 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 as to what works and does not work in regard to 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 5) must be made with caution and must remain speculative because of the relatively large amount of missing data relating to these outcomes.

Discussion

Overall Findings

The most important outcome of the overall effect size analysis relates to the wide variability in outcomes for all three primary measures. While the average effect of DE was near zero, there was a tremendous range of effect sizes (g) in achievement outcomes, from -1.31 to $+1.41$. There were instances in which the DE group outperformed the traditional instruction group by more than 50%, and there were instances in which the opposite occurred, for example, the traditional instructional group outperforming the DE group by 48% or more. Similar results were found for overall attitude and retention outcomes.

None of the measures were homogeneous, so interpreting means as if they are true representations of population values is risky (Hedges & Olkin, 1985). It is simply incorrect to state 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 in regard to 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 be made only 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 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 past few years, a number of commentators (Anglin & Morrison, 2000; Diaz, 2000; Perraton, 2000; Phipps & Merisotis, 1999; Saba, 2000) 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 research literature in terms of its completeness. This discussion begins with that assessment, because both quality of studies and depth of reporting impinge upon all other aspects of the analysis.

One entire 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 nonequivalent 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 affected the accuracy of the methodological assessment, since missing aspects of design, control, measurement, equivalence of conditions, and so forth influence the quality of the assessment.

Information available in the literature. Overall, we found the literature severely wanting in terms of depth of reporting. Nearly 60% of 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 able to derive only a 40% estimate of the effects 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 with a “classroom condition.” If we cannot discern what a DE condition is being compared with, it is very difficult to come to any conclusion as to what is meant by an effect size characterizing differences. This was not just a problem in reports and conference papers, which 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 sources revealed only a small bias in mean effect size among the types of literature 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 in which 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) argued 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 the guidelines of 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).

Since, as mentioned, many codable aspects of methodological quality were unavailable owing to missing information, we attempted to characterize the quality of studies in terms of research design and degree of control for confounding. We chose to enter the 13 methodological study features into a WMR as a way of (a) assessing methodology independently and in relation to other blocks of study features and (b) assessing other study features after variation due to methodology had been removed. We found that methodology accounted for a substantial pro-

portion of the overall variation in 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 few missing data, strong measures, and adequate control over confounding, the variance accounted for by methodology would have been minimal. In the most extreme situation, 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 the 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 as 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. Current use of the term *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 between 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 -1.14 to $+0.97$ (this was the only homogeneous subset), while asynchronous outcomes favored the DE condition, ranging from -1.31 to $+1.41$. While both mean effect sizes for attitudes were negative, the differences were dramatic for synchronous and asynchronous DE, favoring classroom instruction by nearly 0.20 standard deviations. The split for retention outcomes yielded the opposite outcome. Dropout was substantially higher in asynchronous DE than in 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 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 medi-

ated 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 synchronous DE literature 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 of 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. On the basis of 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, and Caswell (2003) claim that this transition can be made fairly easily by experienced instructors. Presumably, this applies to both synchronous and asynchronous DE; however, because synchronous DE is more like classroom instruction and takes place 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, on the basis of 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 reported 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 reside, 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 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—for example, that of Kember (1996)—that stress factors such as entry characteristics, social integration, external attribution, and academic integration.

Only a small percentage of the findings for synchronous DE were 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 with home schooling or the provision of specialized content by a nonexpert (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 with 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 Versus 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 DE 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 had been split into synchronous and asynchronous DE. We settled on an approach to WMR in which blocks of these recoded study features were entered in different orders and assessed the R^2 changes 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 sizes, whether entered in the first or the third step of the 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 used, the feedback provided, and the degree of learner engagement, that create the conditions within 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 in which step in the 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 are not important? No, it cannot 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 DE literature 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 in comparison 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 conferencing, 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, 1985). Second, some of the individual study feature results were 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., problem-based learning [PBL]) that includes (or induces) some degree of 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 telephone) appear to facilitate better attitudes, and “providing advance course information” benefits achievement outcomes.

The results for achievement and attitude across synchronous and asynchronous DE are both strikingly similar and strikingly different. For instance, for asynchronous DE, PBL appears as a strong predictor in favor of the DE condition. Although this was 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), in which 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, and 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. CMC 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. In the case of 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 overarching 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 PBL. 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 fea-

tures 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's or instructor's set of tools. Thus, 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 members of 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 PBL 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, with the 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, wherein superficial learning is promoted through efficient communication (e.g., seeking only the correct answer), and "deep" interaction among learners, wherein 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 sociocognitive and constructivist principles of learning such as those summarized by the American Psychological Association (1997). In addition, learning materials and tasks must engage the learner in ways that promote meaningfulness, understanding, and transfer. Clarity, expressiveness, and feedback may help to ensure learner engagement and interactivity; multimedia learning materials may do likewise when they are 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; DE should not simply be an electronic copy of paper-based material. This may explain why the effect sizes were 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 experiences? These tools need further development and more appropriate deployment. A contrasting view, supported by the size of effects found 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 long-standing speculation that might not be verified until the next generation of DE is widely available and appropriately used. The second alternative requires that policymakers devote energy 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, on 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 will diminish. 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 1950s and 1960s 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 whether there are aspects of design, relating to either medium or method, that 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?” Such comparisons are the stock-in-trade of meta-analysis, but once the basic question is answered, more or less, we should begin to move toward 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 does not 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.”

Thus, if comparison studies do continue—and we suspect that they will—can we envisage an optimal comparative study? In the best of all of 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 this means eliminating all potential confounds—selection, history, materials, and so forth—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 DE and 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 anymore 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 inherent in DE/classroom-comparative research.

The following are some other directions for future research:

- Developing a theoretical framework for the design and analysis of DE. Adapting the learner-centered principles of the American Psychological Association (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 expend less effort learning, especially when they choose between DE and regular courses for convenience purposes (i.e., they are happy to have choice and are 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 in synchronous and asynchronous DE settings might be a place to start. Examining different methods for developing these skill sets might extend from this assessment.
- 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” represents 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 policy-making. 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., CMC) appears on the educational horizon; (c) an assessment of the relative strength and effect of methodological quality on assessments of other contributing factors; (d) a glimpse at 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.

Notes

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 and funding from the Louisiana State University Council on Research Awards and Department. We express appreciation to Janette M. Barrington, Anna Pereti-*atkovicz*, Mike Surkes, Lucie A. Ranger, Claire Feng, Vanikumari Pulendra, Keisha Smith, Alvin Gautreaux, and Venkatraman Kandaswamy for their assistance and contributions. Also, we thank Richard E. Clark, Gary Morrison, and Tom Cobb for their comments on this research and their contributions to the development of ideas for analysis and discussion.

¹We 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. This provided three new composite variables, which were then entered into the WMR in different orders, as indicated earlier. 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 earlier 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.

References

References marked with an asterisk indicate studies included 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*, 151–179.
- Albanese, M. A., & Mitchell, S. (1993). Problem-based learning: A review of literature on its outcomes and implementation issues. *Academic Medicine, 68*, 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. *American Journal of Distance Education, 16*(2), 83–97.
- American Psychological Association, Division 15, Committee on Learner-Centered Teacher Education for the 21st Century. (1997). *Learner-centered psychological principles: Guidelines for teaching educational psychology in teacher education programs*. Washington, DC: Author.
- *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*, 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). Students' reactions to the introduction of videoconferencing for classroom instruction. *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*, 88–91.

- *Bader, M. B., & Roy, S. (1999). Using technology to enhance relationships in interactive television classrooms. *Journal of Education for Business*, 74, 357–362.
- *Barber, W., Clark, H., & McIntyre, E. (2002). Verifying success in distance education. *Proceedings of the World Conference on E-Learning in Corporate, Government, Health, & Higher Education*, 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. (1999, September). *Distance education: Analysis of learning preferences in two sections of undergraduate HBSE-like human growth and development courses: Face-to-face and Web-based distance learning*. Paper presented at the 3rd Annual 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, 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*. Paper presented at the annual conference of the International Visual Literacy Association, Pittsburgh, PA.
- *Beare, P. L. (1989). The comparative effectiveness of videotape, audiotape, and telelecture in delivering continuing teacher education. *American Journal of Distance Education*, 3(2), 57–66.
- Beaudoin, M. (1990). The instructor's changing role in distance education. *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, State University of New Jersey, Newark.
- *Benbunan-Fich, R., Hiltz, S. R., & Turoff, M. (2001). A comparative content analysis of face-to-face vs. ALN-mediated teamwork. In *Proceedings of the 34th Hawaii International Conference on System Sciences*. Retrieved May 14, 2003, from Asynchronous Learning Networks (ALN) database.
- 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, 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, 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, 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. *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. *American Journal of Distance Education*, 16(1), 37–52.

How Does Distance Education Compare With Classroom Instruction?

- *Boulet, M. M., Boudreault, S., & Guerette, L. (1998). Effects of a television distance education course in computer science. *British Journal of Educational Technology*, 29, 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? *American Economic Review*, 92, 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.
- *Bruning, R., Landis, M., Hoffman, E., & Grosskopf, K. (1993). Perspectives on an interactive satellite-based Japanese language course. *American Journal of Distance Education*, 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, 274–288.
- *Burkman, T. A. (1994). *An analysis of the relationship of achievement, attitude, and sociological elements of individual learning styles of students in an interactive television course*. Unpublished doctoral dissertation, Western Michigan University, Kalamazoo.
- *Cahill, D., & Catanzaro, D. (1997). Teaching first-year Spanish on-line. *Calico Journal*, 14(24), 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 annual conference of the American Society for Engineering Education, Seattle, WA.
- Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. New York: Houghton Mifflin.
- *Campbell, M., Floyd, J., & Sheridan, J. B. (2002). Assessment of student performance and attitudes for courses taught online versus onsite. *Journal of Applied Business Research*, 18(2), 45–51.
- *Card, K. A., & Horton, L. (1998, November). *Fostering collaborative learning among students taking higher education administrative courses using computer-mediated communication*. 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://teleeducation.nb.ca/content/pdf/english/DEOSNEWS_11.9_effective-studentoutcomes.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, 81–94.
- Carpenter, C. R., & Greenhill, L. P. (1955). *An investigation of closed-circuit television for teaching university courses* (Report 1). University Park: Pennsylvania State University.
- Carpenter, C. R., & Greenhill, L. P. (1958). *An investigation of closed-circuit television for teaching university courses* (Report 2). University Park: 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, 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, 73–88.
- *Chapman, A. D. (1996). *Using interactive video to teach learning theory to undergraduates: Problems and benefits*. (ERIC Document Reproduction Service No. 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, J., & Armstrong, P. (1991). Comparison of performance and attitude in traditional and computer conferencing classes. *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 teachers' and students' attitudes on a two-way video tele-educational system for Korean elementary school. *Educational Technology Research and Development*, 46, 98–105.
- *Chung, J. (1991). Televised teaching effectiveness: Two case studies. *Educational Technology*, 31, 41–47.
- *Chute, A. G., Balthazar, L. B., & Posten, C. O. (1988). Learning from tele-training. *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 the annual convention of the International Teleconferencing Association, Washington, DC.
- *Cifuentes, L., & Hughey, J. (1998). *Computer conferencing and multiple intelligences: Effects on expository writing*. Paper presented at the annual meeting of the Association for Educational Communication, St. Louis, MO. (ERIC Document Reproduction Service No. ED423830).
- *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, 445–459.
- Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development*, 42(2), 21–29.
- Clark, R. E. (2000). Evaluating distance education: Strategies and cautions. *Quarterly Review of Distance Education*, 1, 3–16.
- *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, LA.
- Cobb, T. (1997). Cognitive efficiency: Toward a revised theory of media. *Educational Technology Research and 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, 353–365.

How Does Distance Education Compare With Classroom Instruction?

- *Collins, M. (2000). Comparing Web correspondence and lecture versions of a second-year nonmajor biology course. *British Journal of Educational Technology*, 31, 21–27.
- Colliver, J. (1999). Effectiveness of PBL curricula. *Medical Education*, 34, 959–960.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- *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.
- *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, 369–386.
- *Curnow, C. K. (2001). *Social interaction, learning styles, and training outcomes: Differences between distance learning and traditional training*. Unpublished doctoral dissertation, 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. In *Proceedings of the Society for Information Technology and Teacher Education International Conference* (pp. 126–130). Charlottesville, VA: Association for the Advancement of Computing in Education.
- *Davis, J. L. (1996, January). Computer-assisted distance learning, Part II: Examination performance of students on and off campus. *Journal of Engineering Education*, pp. 77–82.
- *Davis, R. S., & Mendenhall, R. (1998). *Evaluation comparison of online and classroom instruction for HEPE 129—Fitness and Lifestyle Management Course*. (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. *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=article&id=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, 112–115.

- *Dillon, C. L., Gunawardena, C. N., & Parker, R. (1992). Learner support: The critical link in distance education. *Distance Education, 13*, 29–45.
- Dillon, C. L., & Walsh, S. M. (1992). Faculty: The neglected resource in distance education. *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*, 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*, 454–466.
- *Flaskerud, G. (1994). *The effectiveness of an interactive video network (INV) extension workshop.* University Park: Pennsylvania State University.
- *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 Society for Information Technology and Teacher Education International Conference, 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*, 366–372.
- *Fritz, S., Bek, T. J., & Hall, D. L. (2001). Comparison of campus and distance undergraduate leadership students' attitudes. *Journal of Behavioral and Applied Management, 3*, 3–12.
- *Fulmer, J., Hazzard, M., Jones, S., & Keene, K. (1992). Distance learning: An innovative approach to nursing education. *Journal of Professional Nursing, 8*, 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*, 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. *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. *American Journal of Distance Education, 5*(3), 20–28.

How Does Distance Education Compare With Classroom Instruction?

- *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.
- 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). Bloomington: Indiana University.
- *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*. 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 U\$A" tele-course. *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* (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, 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*. (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, 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 for university level business school instruction. *Journal of Education for Business*, 72, 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* (pp. 24-34). University Park, PA: 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, 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: 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 annual meeting of the American Educational Research Association, New Orleans, LA.
- *Hodge-Hardin, S. (1997). *Interactive television vs. a traditional classroom setting: A comparison of student math achievement*. Retrieved August 23, 2003, from <http://www.mtsu.edu/~itconf/proceed97/hardin.html>
- *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 August 13, 2003, 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*, 400–416.
- *Hurlburt, R. T. (2001). “Lectlets” delivery content at a distance: Introductory statistics as a case study. *Teaching of Psychology, 28*, 15–20.
- *Jeannette, K. J., & Meyer, M. H. (2002). Online learning equals traditional classroom training for master gardeners. *HortTechnology, 12*, 148–156.
- *Jenkins, S. J., & Downs, E. (2002). *Differential characteristics of students in on-line vs. traditional courses*. Retrieved June 4, 2003, from <http://dl.aace.org/10698>
- *Jewett, F. (1998). *The education network of Maine: A case study in the benefits and costs of instructional television*. Seal Beach: California State University, Seal Beach.
- *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*, 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, State University of New York, Buffalo.
- *Johnson, M. (2002). Introductory biology online: Assessing outcomes of two student populations. *Journal of College Science Teaching, 31*, 312–317.
- *Johnson, S. D., Aragon, S. R., Shaik, N., & Palma-Rivas, N. (1999). *Comparative analysis of online vs. face-to-face instruction*. Champaign: 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*(1). Retrieved September 12, 2003, from http://www.usdla.org/html/journal/JAN01_Issue/article01.html
- *Jones, E. R. (1999, February). *A comparison of an all Web-based class to a traditional class*. Paper presented at the meeting of the Society for Information Technology and 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*, pp. 57–60.
- *Kabat, E. J., & Friedel, J. N. (1990). *The Eastern Iowa Community College District's televised interactive education evaluation report*. Clinton: 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*, 19, 15–31.
- *Kahl, T. N., & Cropley, A. J. (1986). Face-to-face versus distance learning: Psychological consequences and practical implications. *Distance Education*, 7, 38–48.
- Kanuka, H., Collett, D., & Caswell, C. (2003). University instructor perceptions of the use of asynchronous text-based discussion in distance courses. *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. *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.
- *Kennedy, R. L., Suter, W. N., & Clowers, R. L. (1997, November). *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, 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, 87–91.
- Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and 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.
- *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*, 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. *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, 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. *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.ictc.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, 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). Sterling, VA: Kogan Page.
- Lou, Y., & MacGregor, S. K. (2002, November). *Enhancing online learning with between group collaboration*. Paper presented at the Teaching Online in Higher Education Online Conference.
- *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, 273–277.
- *Magiera, F. T. (1994–1995). Teaching personal investments via long-distance. *Journal of Educational Technology Systems*, 23, 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 courses. *Behavior 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, 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. *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, 201–211.
- *McCleary, I. D., & Egan, M. W. (1989). Program design and evaluation: Two-way interactive television. *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, 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, 1425–1426.
- *Mills, B. D. (1998). Replication of optimism and pessimism of distance-learning and on campus students. *Psychological Reports*, 83, 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, 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*. (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.
- Morrison, G. R. (1994). The media effects question: “Unresolvable” or asking the right question. *Educational Technology Research and Development*, 42(2), 41–44.
- *Moshinskie, 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.
- *Murray, J. D., & Heil, M. (1987). *Project evaluation: 1986–87 Pennsylvania tele-teaching project*. Mansfield: Mansfield University of Pennsylvania.
- *Muta, H., Kikuta, R., Hamano, T., & Maesako, T. (1997). The effectiveness of low-cost tele-lecturing. *Staff and Educational Development International*, 1, 129–142.
- *Mylona, Z. H. (1999). *Factors affecting enrollment satisfaction and persistence in 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, 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*, 293–302.
- *Neuhauser, C. (2002). Learning style and effectiveness of online and face-to-face instruction. *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*, 285–297.
- Nipper, S. (1989). Third generation distance learning and computer conferencing. In R. Mason & A. Kaye (Eds.), *Mindweave: Communication, computers and distance education* (pp. 63–73). Oxford, England: 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. *New Directions for Teaching and Learning, 1*, 51–58.
- *Parker, D., & Gemino, A. (2001). Inside online learning: Comparing conceptual and technique learning performance in place-based and ALN formats. *Journal of Asynchronous Learning Networks, 5*(2). 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*, 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*, 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*, 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. *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*, 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: Institute for Higher Education Policy.

How Does Distance Education Compare With Classroom Instruction?

- *Piccoli, G., Ahmad, R., & Ives, B. (2001). Web-based virtual learning environments: A research framework and a preliminary assessment of effectiveness in basic IT skills training. *MIS Quarterly*, 25, 401–426.
- *Pirrong, G. D., & Lathen, W. C. (1990). The use of interactive television in business education. *Educational Technology*, 30, 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.
- *Reagan, C. (2002). Teaching research methods online: Course development and comparison to traditional delivery. *Proceedings of the Society for Information Technology and Teacher Education International Conference*, 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., Gabriel, D., & Clegg, A. (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.
- *Ritchie, H., & Newby, T. J. (1989). Classroom lecture/discussion vs. live televised instruction: A comparison of effects on student performance, attitudes, and interaction. *American Journal of Distance Education*, 3(3), 36–45.
- *Rivera, J., & Rice, M. (2002). A comparison of student outcomes and satisfaction between traditional and 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, Alberta, 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: Office of Instructional Telecommunications, University of North Carolina.
- *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.
- *Ryan, W. F. (1996). The effectiveness of traditional vs. audiographics delivery in senior high advanced mathematics courses. *American 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, 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, 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., & 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, 257–265.
- Schoenfeld-Tacher, R., & Perschitte, K. A. (2000). Differential skills and competencies required of faculty teaching distance education courses. *International Journal of Educational Technology*, 2(1), 1–16.
- *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 <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 22, 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*. (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 doctoral dissertation, Touro University International, Cypress, CA.
- Shachar, M., & Neumann, Y. (2003, October). Differences between traditional and distance education academic performances: A meta-analytic approach. *International Review of Research in Open and Distance Education*. Retrieved October 30, 2003, from <http://www.irrod.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.
- Shale, D. (1990). Toward a reconceptualization of distance education. In M. G. Moore (Ed.), *Contemporary issues in American distance education* (pp. 333–343). Oxford, England: 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, 209–231.
- *Simpson, H., Pugh, H. L., & Parchman, S. W. (1993). Empirical comparison of alternative instructional TV technologies. *Distance Education*, 14, 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*

How Does Distance Education Compare With Classroom Instruction?

- tutored video (TVI) and distributed tutored video instruction (DTVI)*. Palo Alto, CA: Sun Microsystems.
- *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, Ann Arbor.
- *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:zSzzSzwwww.compapp.dcu.iezSz~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 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*. Retrieved August 21, 2003, from <http://citeseer.nj.nec.com/cache/papers/cs/20657/http:zSzzSzjimjansen.tripod.comzSzacademiczSzpubszSzszgicse98.pdf/smith98digital.pdfIn>
- *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. *American Journal of Distance Education*, 7(1), 37–53.
- Spector, J. M. (2001). *Competencies for online teaching*. (ERIC Digest Report No. 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, 132–140.
- *Stone, H. R. (1990). *Does interactivity matter in video-based off-campus graduate engineering education?* (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, 212–229.
- *Suter, N. W., & Perry, M. K. (1997, November). *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*. 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, 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: Author.
- *Tiene, D. (1997). Student perspective on distance learning with interactive television. *TechTrends*, 42(1), 41–47.
- *Toussaint, D. (1990). *Fleximode: Within Western Australia TAFE*. (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 Research and 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, 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, Ontario, Canada: 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: 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). 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, 137–143.
- *Waschull, S. B. (2001). The online delivery of psychology courses: Attrition, performance, and evaluation. *Teaching of Psychology*, 28, 143–146.
- *Wegner, S. B., Holloway, K. C., & Garton, E. M. (1999). The effects of Internet based instruction on student learning. *Journal of Asynchronous Learning Networks*, 3(2). Retrieved May 31, 2002, from http://www.aln.org/alnweb/journal/Vol3_issue2/Wegner.htm
- *Westbrook, T. S. (1998). Changes in students' attitude toward graduate business instruction via interactive television. *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.

How Does Distance Education Compare With Classroom Instruction?

- *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, Minneapolis.
- *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 cyborg writing. *Computers and the Humanities*, 29, 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? In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp. 1161–1165). Santa Monica, CA: Human Factors and Ergonomics Society.
- *Wisher, R. A., & Priest, A. N. (1998). Cost-effectiveness of audio teletraining for the U.S. Army National Guard. *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. *American Journal of Distance Education*, 15(3), 20–23.

Authors

ROBERT M. BERNARD is Professor of Education at Concordia University, Montreal, Quebec, Canada, and a member of the Centre for the Study of Learning and Performance, LB-545-5, Department of Education, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; e-mail bernard@education.concordia.ca. He specializes in instructional technology, distance education, online teaching and learning, research design and statistics, and research synthesis (meta-analysis).

PHILIP C. ABRAMI is Professor of Education at Concordia University and Director of the Centre for the Study of Learning and Performance, LB-545-5, Department of Education, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; e-mail abrami@education.concordia.ca. His areas of expertise include instructional technology, social psychology of education, postsecondary instruction, and research synthesis (meta-analysis).

YIPING LOU is Assistant Professor of Educational Technology in the Department of Educational Leadership, Research, and Counseling, Louisiana State University, 111 Peabody Hall, Baton Rouge, LA 70803; e-mail ylo@lsu.edu. She specializes in technology-mediated instruction, collaborative learning, and meta-analysis.

EVGUENI BOROKHOVSKI is a doctoral candidate in the Department of Psychology, Concordia University, and a research assistant at the Centre for the Study of Learning and Performance, LB-545-5, Department of Education, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; e-mail eborokhovski@education.concordia.ca. He specializes in systematic reviews in the area of cognitive psychology and learning (e.g., early reading acquisition, second-language learning).

ANNE WADE is Manager and Information Specialist at the Centre for the Study of Learning and Performance, LB-545-5, Department of Education, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; e-mail wada@education.concordia.ca. Her expertise is in information storage and retrieval and research strategies.

LORI WOZNEY is a doctoral candidate in Concordia University's Educational Technology program and is a member of the Centre for the Study of Learning and Performance, LB-545-5, Department of Education, Concordia University, 1455 de Maisonneuve Blvd. West,

Bernard et al.

Montreal, Quebec, Canada H3G 1M8; e-mail wozney@education.concordia.ca. Her work focuses on the integration of instructional technology, self-regulated learning, and organizational analysis.

PETER ANDREW WALLET is an MA student in educational technology and a research assistant at the Centre for the Study of Learning and Performance, LB-545-5, Department of Education, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; e-mail wallet@education.concordia.ca. Teacher training using distance education is a prime interest in his research.

MANON FISET is an MA student in educational technology and a research assistant at the Centre for the Study of Learning and Performance, LB-545-5, Department of Education, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; e-mail fiset@education.concordia.ca. She has worked in the field of distance education for several years at the Institute of Canadian Bankers and the Centre for the Study of Learning and Performance.

BINRU HUANG is a research assistant in the Department of Educational Leadership, Research, and Counseling, Louisiana State University, 111 Peabody Hall, Baton Rouge, LA 70803; e-mail binruhuang@hotmail.com. Her research interests include statistical analysis methods and use of technology in distance learning.

APPENDIX A

Coded variables and study features: 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 toward course
 4. Attitude toward the technology
 5. Attitude toward the subject matter
 6. Attitude toward 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

(continued)

APPENDIX A

(Continued)

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 "instructional design" (Name: "Id")*
3. DE condition: Advance information (Name: "Advinf")
 1. Information received before commencement of course
 2. Information received at first course
 3. No information received
4. Opportunity for face-to-face contact with instructor (Name: "f2ft")
 1. Opportunity to meet instructor during instruction
 2. No opportunity to meet instructor
 3. Opportunity to meet instructor prior to, or at commencement of, instruction only (e.g., orientation session)
5. Opportunity for face-to-face contact with peers (Name: "f2fp")
 1. Opportunity to meet peers during instruction
 2. No opportunity to meet peers
 3. Opportunity to meet peers at or prior to commencement of instruction
6. Provision for synchronous technically mediated communication with teacher (Name: "Syncte")
 1. Opportunity for synchronous communication
 2. No opportunity for synchronous communication
7. Provision for synchronous technically mediated communication with 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")*

APPENDIX A

(Continued)

Section F: Media Features

1. Use of two-way audio conferencing (Name: "Ac")*
2. Use of two-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 one-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 of 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
3. Instructor experience with DE (Name: "Inde")
 1. Yes
 2. No
4. Instructor experience with technologies used (Name: "Intech")
 1. Yes
 2. No
5. Student experience with DE (Name: "Stude")
 1. Yes
 2. No
6. Student experience with technologies used (Name: "Stutech")
 1. Yes
 2. No
7. Types of control learners (Name: "Lrtypc")
 1. K-12
 2. Undergraduate
 3. Graduate
 4. Military
 5. Industry/business
 6. Professionals (e.g., doctors)
8. Types of DE learners (Name: "Lrtypd")
 1. K-12
 2. Undergraduate
 3. Graduate
 4. Military
 5. Industry/business
 6. Professionals (e.g., doctors)
9. Setting (Name: "Settg")
 1. DE urban and control rural
 2. DE urban and control urban
 3. DE reported/control not reported

(continued)

APPENDIX A

(Continued)

-
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 statistics and algebra)
 2. Languages (includes language arts and second languages)
 3. Science (including biology, sociology, psychology, and 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
 11. Average age (Name: "Age")
 1. Real difference in age means with corresponding sign
-

Note. Items followed by asterisks were coded according to 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)

APPENDIX B

Categories, numbers, and percentages of excluded studies

Category	Excluded studies	
	No.	%
Review and conceptual articles	52	8.25
Case studies, survey results, and qualitative studies	55	8.73
Studies with violations of either DE or face-to-face definitions	295	46.83
Collapsed data, mixed conditions, or program-based findings	43	6.83
Insufficient statistical data	97	15.40
Nonretrievable studies	10	1.58
"Out-of-date" studies	21	3.33
Duplicates	11	1.75
Multiple reasons	46	7.30
Total	630	100

APPENDIX C

Dates and categories of publication for achievement outcomes

Publication date category	Frequency	%	
1985–1989	27	8.49	
1990–1994	91	28.61	
1995–1999	108	33.96	
2000–2002	92	28.93	

Publication category	Frequency	Relative %	g+
Journal articles	135	42.45	–0.009
Dissertations	64	20.13	0.022
Technical reports	119	37.42	0.036*

* $p < .05$.

APPENDIX D

Effect sizes for demographic study features ($k \geq 10$)

Study feature	g+	t
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**
Types 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**

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