What Makes the Difference? A Practical Analysis of Research on the Effectiveness of Distance Education

YONG ZHAO
Michigan State University

JING LEI
Syracuse University

BO YAN
CHUN LAI
Michigan State University

HUEYSHAN SOPHIA TAN
Coastal South Carolina University

This article reports findings of a meta-analytical study of research on distance education. The purpose of this study was to identify factors that affect the effectiveness of distance education. The results show that although the aggregated data of available studies show no significant difference in outcomes between distance education and face-to-face education as previous research reviews suggest, there is remarkable difference across the studies. Further examination of the difference reveals that distance education programs, just like traditional education programs, vary a great deal in their outcomes, and the outcome of distance education is associated with a number of pedagogical and technological factors. This study led to some important data-driven suggestions for and about distance education.

> History is a mirror of the past and a lesson for the present.
> —Persian proverb

Research in distance education[1] has traditionally been dominated by what is often referred to as "comparison studies" or "media comparison studies" (R. E. Clark, 1983; R. E. Clark & Salomon, 1986; Gunawardena & McIsaac,

Teachers College Record Volume 107, Number 8, August 2005, pp. 1836–1884

0161-4681

2004; Lockee, Burton, & Cross, 1999; McIsaac & Gunawardena, 1996). Typically these studies compared the effectiveness of distance education with that of face-to-face education or the effectiveness of one technology over another (Gunawardena & McIsaac, 2004; Joy & Garcia, 2000; Lockee et al., 1999; McIsaac & Gunawardena, 1996). Most of these studies found no significant differences in learning outcomes (Gunawardena & McIsaac, 2004; McIsaac & Gunawardena, 1996; Russell, 1999). Although some studies found positive effects, their effects are then quickly balanced out by other studies that found negative effects (Gunawardena & McIsaac, 2004; McIsaac & Gunawardena, 1996). Thus as a whole, distance education has been said to be as effective as its face-to-face counterpart (Institute for Higher Education Policy, 1999).

These comparison studies have been seriously criticized for many reasons: focusing on the wrong factor, methodologically flawed, biased sampling, using improper measures of outcomes, even being pseudoscientific (R. E. Clark, 1983; R. E. Clark & Salomon, 1986; Gunawardena & McIsaac, 2004; Joy & Garcia, 2000; Lockee et al., 1999; McIsaac & Gunawardena, 1996). But the most serious charge is that these studies are futile and useless. "Whatever methods have been used to report the results of media comparison studies and their instructional impact, these studies have yielded very little useful guidance for distance education practice" (Gunawardena & McIsaac, 2004, p. 378). On these grounds, researchers have been called to discontinue this line of research (R. E. Clark, 1994; Gunawardena & McIsaac, 2004; Lockee et al., 1999; McIsaac & Gunawardena, 1996). Instead, they are advised to move "beyond the no-significant difference" (Twigg, 2001, p. 7), to use different designs, and to focus on different factors (R. E. Clark, 1994; Gunawardena & McIsaac, 2004; Koumi, 1994; Lockee et al., 1999; McIsaac & Gunawardena, 1996).

Although the call for a new paradigm of research in distance education has not stopped researchers from conducting comparison studies in distance education, as publications of this nature continue to appear in scholarly journals and professional conferences (Lockee et al., 1999), the characterization of previous comparison studies as flawed and useless has the potential, if it has not already done so, to lead to a complete dismissal of a large body of literature. And that could be a mistake, because these studies may have what we are looking for—useful guidance for future practices and research—when examined with a different lens.

Traditionally, the lens used to synthesize media comparison studies has often in essence been the same as the one used by comparison studies: the difference finder. A number of synthesis studies have been conducted in the field of distance education (Allen, Bourhis, Burrell, & Mabry, 2002; Cavanaugh, 2001; Dubin & Hedley, 1969; Machtmes & Asher, 2000; M. G. Moore & Thompson, 1996; Schlosser & Anderson, 1994; Schramm, 1962). These studies, not unlike the studies they synthesize, attempted to assess whether distance education or its delivery media, taken as a whole, produced better, worse, or the same learning outcomes relative to face-to-face instruction. While some of these review studies, especially earlier ones, did qualitative analysis of the literature, a few more recent ones employed the meta-analysis method when more experimental or quasi-experimental studies became available. Regardless of the approach, these studies resulted in complaining either that the literature has too many methodological problems to lead to any conclusion (Schlosser & Anderson, 1994; Stickell, 1963) or that there is no significant difference between distance and face-to-face instruction (Cavanaugh, 2001; Machtmes & Asher, 2000; M. G. Moore & Thompson, 1996; Russell, 1999).

These studies have undoubtedly helped to promote distance instruction as a viable form of education with the same quality as its face-to-face counterpart; they have also led to the belief that previous distance education research is of low quality and has little to offer in terms of practical guidance for improving practice. Consequently, it is suggested that we should just discard it and move forward. However, upon closer examination, these studies reveal that individual studies indeed found significant difference between distance and face-to-face instruction. It is likely that a systematic analysis of what may account for the different findings across studies could provide us with practical guidance for improving practice.

The no-significant-difference conclusion was primarily drawn from two types of analyses: summary of studies that found no significant difference (Russell, 1999) and meta-analysis (Cavanaugh, 2001; Machtmes & Asher, 2000). The most influential and representative of the first type is Russell's inventory of studies that found no significant difference. In 1999, Russell published an annotated bibliography of studies that found no significant difference between face-to-face and distance education. These studies span almost a century, beginning in 1928 and ending in 1998. This impressive collection of 355 articles forcefully supported the *no-significant-difference phenomenon,* a term used by Russell to refer to effectiveness studies of distance education (including instruction using technology). However, most of the studies on this widely circulated list were not experimental studies. Instead, they were "surveys with small sample sizes (less than forty), no mention of the return rate of the surveys, and no mention of the learner demographics" (Machtmes & Asher, 2000, p. 31). Furthermore, these studies were not identified using any systematic approach. According to Russell: "I did not use any scientific sampling method but instead listed every study found that showed no significant difference. . . . The point remains that such studies are practically nonexistent and the very few that do exist are offset by a like number which show negative results for the technology-based instruction" (p. xiii). For these reasons, the validity and reliability of Russell's claim should be questioned.

A more valid and reliable way to synthesize the literature is meta-analysis (Glass, 1976; Glass, McGaw, & Smith, 1981) because of its systematic procedure and criteria for identifying and selecting previous studies and verified statistical procedures for analyzing results. Interestingly, the two meta-analyses available (Cavanaugh, 2001; Machtmes & Asher, 2000) came to essentially the same conclusion as Russell: There is no significant difference between the outcomes of distance and face-to-face education. However, these two meta-analyses also found considerable differences among studies in terms of effect size, a measure of difference in learning outcomes between the two comparison groups, in this case, distance education and face-to-face education. For example, Cavanaugh (2001), after calculating the effect sizes of 19 studies that compared K–12 students learning with interactive distance education technology with students learning with traditional classroom instruction, found the weighted mean effect size across all studies to be 0.147, but the standard deviation was 0.69, indicating a significant variation among the studies. In other words, these studies were very different from each other in terms of their findings. Machtmes and Asher found the same phenomenon: There exists tremendous variation in the outcomes of distance education and face-to-face education. The overall effect size of the 19 studies comparing distance and traditional classroom instruction for adults they analyzed was -0.0093, with a range of -0.005 to $+1.50$. They found that "considerable heterogeneity was indicated ($H = 47.927$, $df = 18$, $p = 0.0002$)" (p. 36).

The considerable heterogeneity in the studies clearly indicates that there is indeed significant difference in learning outcomes of distance and face-to-face education. Individually many studies found significant differences between distance and face-to-face education, some favoring distance education and others face-to-face education. In fact, contrary to Russell's claim, it is rarely the case that the individual studies included in the meta-analyses conducted by Cavanaugh (2001) and Machtmes and Asher (2000), which are ostensibly of higher quality than Russell's, reported no significant difference between distance and face-to-face instruction. However, the difference disappears when the studies are considered as a whole.

The significant heterogeneity of achievement begs the question: Why did some studies find distance education students had better achievement than their counterparts in traditional classrooms and some find the opposite? To answer this question, we need to further examine the characteristics of each individual study. We know that distance education programs vary a great deal in content, learner characteristics, instructor characteristics, and delivery method. By examining these variables and the degree to which they influence learning outcomes, we may be able to arrive at what distance education research is encouraged to do: find useful guidance for practice and research.

Machtmes and Asher (2000) made an attempt to do so as part of their meta-analysis of the effectiveness of telecourses in distance education. They coded 23 contextual variables to describe the qualities of each study. These variables were grouped into two categories of features of the studies: instructional features (e.g., course type, type of delivery equipment, instructor's experience with delivery method) and methodological features (e.g., decade of study, research design type, and methods of assessing achievement). Each of these features was examined to see whether and to what extent it could explain the heterogeneity in the effect sizes. Among other things, Machtmes and Asher found that studies that employed two-way interaction technology were "the only type [of delivery] that had a positive effect size" (p. 38) and that "learner achievement was influenced both by the type of course offered and by the type of learning environment" (p. 40). Another interesting finding is that the time when the study was conducted has a large impact on learner achievement. Similarly, although Cavanaugh (2001) did not attempt deliberately to explain the heterogeneity found in her meta-analysis, her grouping the studies into different content areas did suggest that some groups of studies seem to have larger effect sizes than others, suggesting that some content may be more suitable for distance instruction.

Machtmes and Asher's study suggests a promising way to make use of the distance education literature, but it has a number of significant limitations. First, it has an extremely small sample. With only 19 studies from 13 publications that span 30 years (from 1963 to 1993), the power of the study is extremely limited. Second, the study is limited to video-based/televised distance programs, whereas in recent years distance education has employed many other technologies, including computer conferencing, the World Wide Web, and CD-ROMs. "In order to identify which features impact student learning," they suggest, "researchers need to systematically identify and evaluate the technological and instructional features of all delivery systems" (p. 42). Third, the coding of the variables in the study is all categorical, whereas we know that some of the features vary on a continuum. For example, the availability of the instructor is not a simple yes or no, because some distance education programs may have instructors available more frequently than others. Thus treating it as a continuous variable more accurately reflects the reality. The last but perhaps most significant limitation is the lack of a well-developed framework for identifying possible features that may contribute to learner achievement.

The purpose of the present study is in some way a continuation of Machtmes and Asher's, but with more emphasis on systematically examining how different features of previous studies of distance education affect learning outcomes so as to inform future practice and research. To avoid the limitations of previous synthesis studies, we examined a much larger body of research, applied more sophisticated statistical procedures, and developed a more systematic analytical framework. In the remainder of this article, we describe the methods, findings, and implications of the present study.

METHODOLOGY

LITERATURE SEARCH AND SELECTION

Studies included in the research synthesis were identified through a three-step process. First, we conducted a thorough search for all studies included in the Education Resources Information Center (1966–2002) through FirstSearch with the following keywords: distan* and education, distan* and learning, distan* and teaching, distan* and instruction, online and education, online and learning, online and teaching, online and instruction, on-line and education, on-line and learning, on-line and teaching, on-line and instruction, web-based and education, web-based and learning, web-based and teaching, web-based and instruction, virtual and education, virtual and learning, virtual and teaching, virtual and instruction. The search identified 8,840 potentially relevant articles. Citation information for all 8,840 articles was then transferred into EndNote (version 5.0; ISI ResearchSoft, 2001) to build the first database.

At the second stage, the database was further examined based on the following criteria:

1. The article had to be published in a journal. The decision to include only journal articles was based on the concern of study quality. Previous research reviews on distance education had pointed out the low-quality problem of most studies. We believed that journal articles were of higher quality because of peer review procedures. Only including journal articles may result in publication bias, but we believed that the risk was minimal, as there had not been a dominant paradigm for distance education over the years to cause a certain bias against or for positive, negative, or nonsignificant findings.

2. The article must have had complete reference information (author, date, source, etc.).

3. The article had to include at least one evaluation study of distance education. The specific outcome measured was not limited.

4. The article must have had at least one comparison study on distance education and face-to-face education. Studies in which students' own pretreatment scores served as controls for their posttreatment scores and those in which one distance course was compared with another distance course were excluded.

5. The article must have had some empirical data about the learning outcomes. Articles were not included if they merely describe a distance education course.

6. The article had to include enough statistical information for computing an effect size. The specific information we were looking for was mean, standard deviation, and sample size for both the distance education group and the face-to-face group, or t value, F value and degree of freedom (df).

A total of 1,100 articles that either were not journal articles or didn't have complete reference information were removed from the database. Then the research team read the abstracts of the remaining 7,740 articles, 6,365 of which were removed because they didn't meet criterion 3 or 4. Articles on which a clear decision could not be reached at this stage were kept in the database to be dealt with at a later stage. A total of 1,375 references were left after this selection.

At the third stage, the research team collected and read the 1,375 articles and excluded those that didn't have empirical data. As a result, 421 articles were left after this elimination. The research team then read and coded these 421 articles and found that only 49 articles contained sufficient information for calculating the effect size. For fear of missing some articles that might actually have had the information because of the large number of articles and complexity of the database, the research team examined the 421 articles once more and identified 2 more articles that had complete information for calculating the effect size. Ancestry search was conducted, but no extra articles that met all the criteria were found. Thus, 51 journal articles were included for the analysis.

ANALYTICAL FRAMEWORK

To identify those methodological and substantive characteristics that may be responsible for significant variations in the findings, a detailed analytical framework was developed through an iterative process. The framework was developed based on our understanding of possible sources of variation in the studies. Heterogeneity in outcomes across studies can come from three sources: the publication, the study, and the instruction. Figure 1 depicts the logic model underlying the analytical framework.

For publication and study features, we started with Stock's (1994) seven categories for describing research reports: report identification, the setting of the study, participants, methodology, treatment characteristics, statistical outcomes or effect sizes, and coding process. Instructional features refer to the characteristics of the distance education program under study. It has been argued that distance education should be considered as education at a distance (Shale, 1990):

> In sum, distance education ought to be regarded as education at a distance. All of what constitutes the process of education when teacher

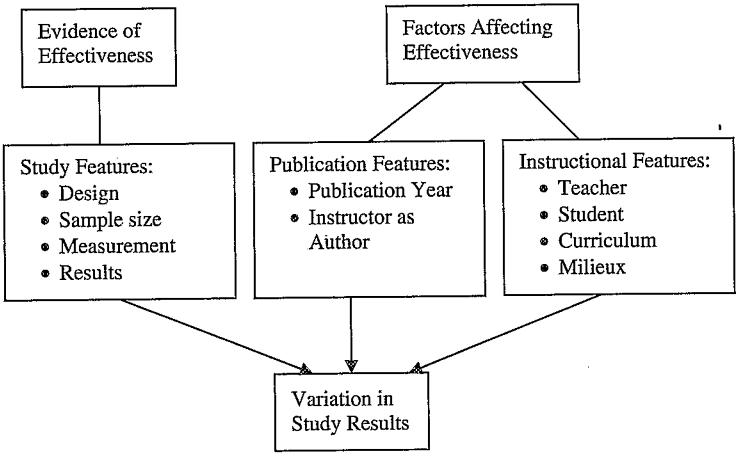

Figure 1. Logic Model of Distance Education Effectiveness

> and student are able to meet face-to-face also constitutes the process of education when teacher and student are physically separated. (p. 334)

In other words, the quality of distance education programs is influenced by the same set of factors that affect the quality of face-to-face education. Schwab (1983) characterizes education in terms of four common places of education: teacher, student, what is taught, and milieux of teaching-learning. This characterization is also applicable to the study of distance education. Whereas teacher, student, and what is taught remain pretty much the same as in face-to-face education, the milieux of teaching-learning are different between distance and face-to-face education in that the milieux of teaching-learning of distance education are mostly mediated through some kind of technology. Hence we describe the milieux of teaching-learning in terms of the format and method of delivery. Finally, we used the grounded theory (Glaser, 1992; Glaser & Strauss, 1967) to guide the development of the framework. Following a typical constant-comparison process, we started coding the studies with the initial framework and then modified the framework when we came across new features during the coding process. Table 1 summarizes the variables included in the final framework. In the following paragraphs, we describe these variables and reasons for their inclusion in the framework.

Evidence of Effectiveness

There are different ways to measure the effectiveness of distance education programs. Studies in distance education thus differ in what they used as

Table 1. Variables included in the study

Category	Variable	Scale	Source
Publication features	Publication year		Publication
	Instructor as author	Yes/no	Publication
Study features	Study design	Quasi-experimental or experimental	Researcher coding
	Measurement	Standardized test, researcher-developed instrument, instructor-developed instrument, and published instrument	Publication
	Indicators of effectiveness	Grades, student satisfaction, faculty satisfaction, standardized tests	Publication
	Results	Sample size, mean, standard deviation, *t* value, *p* value	Publication
Instructional features: instructor	Instructor involvement	1 (*No involvement*) to 10 (*Complete involvement*)	Researcher coding
	Status	Professor, graduate student, or professional	Publication
Instructional features: learner	Background	High school diploma or college degree	Publication
	Status	Full time or part time	Publication
Instructional features: curriculum	Content area	Subject taught	Publication
	Degree	Yes/no	Publication
	Credit	Yes/no	Publication
	Course time		Publication
	Class time		Publication
Instructional features: milieux	Interaction type	Asynchronous interaction, synchronous interaction, both, or noninteractive involvement	Publication
	Media involvement	1 (*No involvement*) to 10 (*Complete involvement*)	Researcher coding
	Setting	K–12, graduate, undergraduate, military, etc.	Publication

evidence of effectiveness and the reliability and validity of the evidence used. The variation in what was measured and the quality of the measurement may explain the heterogeneity of outcomes.

Outcome measures. Information about what has been used to assess the effectiveness of distance programs was collected for each study. A study could use one or more of the following measures: grades, quizzes, independent/standardized tests, student satisfaction, instructor satisfaction, dropout rate, student evaluation of learning, student evaluation of course, and external evaluation. Grades usually are the final scores students received for the class. Student evaluation of learning is students' perception of how much they learned from the course, which can be significantly different from the grades they received.

Source of instrument. The source of instruments used to measure effectiveness can affect the final outcomes in that instruments from different sources may have different levels of reliability and validity. We identified four sources of the most frequently used measures: commercial testing agencies, the researcher (the author of the article), the instructor of the course, and publishers of textbooks that include assessment items.

Study design. The design of a study is a good indication of its quality and thus the quality of the results. We were curious about whether a certain type of design is associated with the study results; thus we coded the studies into two categories: true experimental or quasi-experimental. The differentiating characteristic between the two designs is whether random sampling method was used.

Study results. To calculate effect sizes, the results of each study were recorded in the database. The study results include means, standard deviations, t values, F values, and r values, depending on what is reported in the primary study.

Factors Affecting Effectiveness

Factors that may affect the effectiveness of a distance education program can be categorized into two groups: publication features and instructional features. Two publication features were identified as possible factors affecting effectiveness: publication year and instructor as author.

Publication year. Previous research found that the time when a study was conducted had a significant correlation with the reported effectiveness (Machtmes & Asher, 2000). The assumption was that technology used to deliver distance education had changed dramatically over the years and that newer technologies seemed to have more capacity to deliver richer and more powerful learning experiences. To verify this hypothesis, we coded the year when a study was published for each article.

Instructor as author. All studies are based on advocacy (Begg, 1994). We hypothesize that if the instructor of the distance learning course was also the author of the publication, the result could be more likely to favor distance learning. To verify this hypothesis, we coded whether the author of an article was also the instructor of the distance education program under study.

As mentioned before we used Schwab's four common places of education to guide the identification of instructional features that could potentially affect the effectiveness of distance education programs: the teacher, the student, the curriculum, and the milieux. In each of the four common places were a number of potential factors that could contribute to the outcomes of learning.

Teacher: Instructor involvement. The extent to which the instructor of a distance education course is involved in the actual delivery of the content and available for interactions with students during and outside the class sessions is termed "instructor involvement." The level of instructor involvement is perhaps one of the most defining differences between traditional face-to-face education and distance education. In face-to-face education, the instructor generally delivers the content live and interacts with students both in and outside class meetings, whereas in distance education programs the level of instructor involvement varies a great deal, from one extreme where the content is preprogrammed and delivered through some technology means without the actual involvement of an instructor to another where the instructor actually delivers the content live and is available for interactions with students in very much the same fashion as face-to-face education. Interactions between the teacher and students have been found to affect the quality of student experiences and learning outcomes in distance education (Institute for Higher Education Policy, 2000). How content is delivered should also have an effect on student learning experiences and outcomes. This is also a hot topic for distance education programs because one of the appeals of distance education is the potential to increase efficiency by reducing the demand of actual involvement of faculty. Having one faculty member teaching thousands of students (or even more) with the help of broadcasting, recording, and computing technologies has been a dream for many advocates of distance education. However, if a higher level of instructor involvement becomes a requirement for effective programs, the efficiency dream may never be realized. On the other hand, if the level of instructor involvement is found to be irrelevant to student outcomes, it would be unnecessary to assign an instructor to only a small group of students or have him or her be actually involved in the teaching. A preprogrammed video or computer program can accomplish as much. To assess the value of instructor involvement, we hence included this factor. We coded instructor involvement on a scale from 1 (*no human involvement;* e.g., computer-based training) to 10 (*full involvement of a human instructor;* e.g., two-way interactive TV courses).

Teacher: Status of the instructor. Another way distance education programs have sought to increase efficiency is to employ nonregular faculty, who normally cost less than regular faculty. Thus we were interested in whether instructor status influenced student learning in distance education programs.

Teacher: Training for teaching distance courses. It has been argued that instructors of distance education programs should be trained first, because distance education is a different teaching environment from face-to-face classrooms. Again, the training has cost implications for programs. To test whether training affects student learning, we collected information about teacher training from each study, if that information was available.

Student: Education level. We collected information about students' educational attainment level before attending the distance course to examine whether, and if so at which level, certain types of students were more prepared to take distance education courses.

What is being taught: Content area. Some content may be more suited for distance education, whereas other content may be better taught in a face-to-face course. Interested in whether this assumption was true and if so, what content area was better suited for distance education, we collected information about the content area of each study. A course could be categorized as teaching one of the following subject areas: social science, mathematics, science, medical science, literacy, humanities, business, law, engineering, computer science, teacher education, and skills. (Skills here represented any professional training that didn't fall into other categories.) We coded medical science, business education, and teacher education separately because they had been among the most commonly taught content areas in distance education.

The milieux: Instructional level. Distance education programs have been traditionally intended for adults, but recently distance education has expanded to include younger audiences. As a related factor to student characteristics and content, the instructional level of distance education may be associated with its effectiveness. We grouped the distance education programs in each study into nine levels: Grades K–2 (lower elementary), Grades 3–5 (upper elementary), Grades 6–9 (middle school), Grades 10–12 (high school), associate's degree (community college), undergraduate level (4-year college), graduate level, professional development, and military training. We also collected data about whether the course was for credit or not and whether it was for a degree-granting program or not.

The milieux: Interaction type. Interaction type characterizes how instructors and students interact in the distance learning process. There are four types of interaction: asynchronous, in which a time lag exists between the interactions of the instructor and students in that students may ask a question via e-mail, to which the instructor may respond, for example, 2 days later; synchronous, where the potential exists for instructors and students to interact at the same time; noninteractive, where there is no interaction between instructors and students at all; and both synchronous and asynchronous, where the instructor can interact with students both synchronously and asynchronously.

The milieux: Media involvement. Distance education programs also vary in the level of technology used. Some programs employ a mixed model, in which part of the instruction is conducted face-to-face whereas some others are delivered via technology. Proponents of the mixed model suggest that some face-to-face contact is necessary or desirable to maintain student motivation and thus a higher quality of education. We were interested in testing this hypothesis. Thus we coded each study's level of media involvement, which was defined as the extent to which a certain instructional delivery system has been mediated by technologies, that is, how frequently technology is used in a program. Media involvement is coded on a scale from 1 (*no technology was used*) to 10 (*instruction was delivered completely with technology*).

DATA CODING PROCESS AND INTERRATER RELIABILITY

Information from complete articles selected for inclusion was coded by the two researchers who were most involved in the development of the framework and rubric. One coded 25 articles and the other coded 26 articles independently. When there was any uncertainty, a third researcher was involved, and an agreement would be reached through discussion. After both researchers completed the coding, 10 articles were randomly selected to test interrater reliability. Both coders coded these 10 articles, and they reached an agreement of 98.3%. Disagreements were solved through discussion.

DATA INTEGRATION

Effect Size Computation

Effect size is a measure of standardized mean difference between two groups. In this study, effect size was computed to estimate the extent of the difference between online learning and face-to-face learning. Depending on the information available, we used different strategies to compute the effect size for each study. When information on mean and standard deviation for both control and experimental groups was available, effect size was computed by subtracting the control group mean (face-to-face education) from the experimental group (distance education) mean and dividing the difference by their pooled standard deviation.[2] When information on means and standard deviations was not available and only t values were reported, the effect size was computed based on t value and degrees of freedom.[3] When only the F values and sample sizes were reported, and there were only two groups, the effect size was computed based on the F value and sample sizes (see Rosenthal, 1991).[4] Positive effect sizes indicate that distance education has a better outcome than face-to-face education and vice versa.

Correcting Sample Size Bias

One statistical principle is that studies with larger within-study sample sizes will give more accurate estimates of population parameters than studies with smaller sample sizes (Shadish & Haddock, 1994). In meta-analysis, effect sizes are biased in studies with smaller sample sizes. Hence, large-sample studies should be weighted more than studies with smaller sample sizes. In this study, all effect sizes were corrected for potential bias introduced by different sample sizes.[5]

Multiple Outcomes

In some cases, one primary study has more than one outcome. Cooper (1998) pointed out that multiple results could happen for two reasons. First, more than one measure of the same construct might have been employed and each measure analyzed separately. Second, different samples of the same population might be used in the same study and their data analyzed separately. Additionally, different times of measurement can also result in multiple outcomes (Lipsey, 1994, p. 112). These multiple results from the same study can be problematic for meta-analysis because the separate estimates in the same study are not completely independent: They share historical and situational influences, and some of them even share influences contributed by having been collected from the same people (Cooper, 1998). Becker (2000) proposed several ways to address the problem of multiple outcomes. Based on her suggestion, we addressed the multiple-outcomes problem in a number of different ways, depending on the specifics of the study. First, an average effect size was derived when the same construct with the same sample was measured with more than one instrument, the same construct was measured by more than one subconstructs, or the same data were analyzed using different statistical methods. Second, one "best" effect size was chosen in the following situations: (a) If the primary study provided comparisons of control groups and experimental groups and comparisons of subsamples, such as at different grade levels, then only the comparison of the whole-sample groups was used to compute the effect size, and (b) If the same construct with the same sample was measured at different time points, one effect size was calculated based on data from the final time point (the end of the course), since we were comparing the effectiveness of the whole course. However, when different constructs were measured, or the same construct was measured on different samples, the effect size for each construct or sample was calculated and included in this study.

The analysis resulted in 99 effect sizes. To examine the impact of extreme values on the data set, outlier analyses were conducted using standard residual procedure and three outliers with standardized residuals larger than $\pm$ 2.0 were identified. Further examination of the outliers found that two of them were due to computational error, and these were then corrected. There was no computational error for the other outlier, which was larger than 3 (3.08). This effect size was deleted from the data set. Consequently, 98 effect sizes from 51 studies on 11,477 participants (8,660 independent participants) were left. The articles and their study features are listed in Appendix A. Strictly speaking, since there were still multiple outcomes from the same studies, the overall Q—the amount of total variance—of the effect sizes could be problematic. However, in later analyses, these effect sizes were categorized into different groups and analyzed separately, thus in effect, the multiple-outcomes problem should not have affected the final results.

PUBLICATION BIAS AND REPRESENTATIVENESS OF THE DATA SET

We examined the representativeness of our data set by checking its distribution and publication bias. The stem-and-leaf plot (Figure 2) demonstrates the distribution of our data set. Statistically, if there is no publication bias, the plot should be symmetric and normally distributed. As shown in Figure 2, the typical values in the stems corresponding to values of d in a range falling between 0.01 and 0.09. The distribution was symmetric and seems to have more negative values. There was no gap in this distribution, indicating no atyptical observations. These features of this stem-and-leaf plot suggest that the effect sizes in this data set were normally distributed.

Publication bias is the tendency on the parts of investigators, reviewers, and editors to submit or accept manuscripts for publication based on the direction or strength of the study findings (Dickersin, 1990). Publication bias has been one of the major concerns in meta-analysis, because a data set with publication bias will lead to biased conclusions (Begg, 1994). Since

```
Stem Leaf                       #
  14 3                          1
  12 3                          1
  10 4                          1
   8 4494                       4
   6 16727                      5
   4 06937                      5
   2 014556913466              12
   0 357115579                  9
  -0 966644331999888775332     21
  -2 7320008763300             13
  -4 832188640                  9
  -6 82953                      5
  -8 7183                       4
 -10 70840                      5
 -12 1                          1
 -14 84                         2
     ----+----+----+----+-
Multiply Stem.Leaf by 10**-1
```

Figure 2. Stem-and-Leaf Plot of the Data Set

publication bias is inversely related to sample size (Begg, 1994), to examine publication bias of the data set in this study, a funnel plot (Figure 3) was generated. A funnel plot is a graph of sample size versus effect size. If there is no publication bias, the distribution should resemble an inverted funnel. This plot is symmetric and looks like a funnel, so we assumed that there was no publication bias in our data set.

The stem-and-leaf plot and publication bias plot showed that the effect sizes in our data set were normally distributed and there was no publication bias. Hence the primary studies included in this study can be said to be representative of the literature in distance education.

ANALYSES

We tested the homogeneity of the effect sizes with the homogeneity analysis (Cooper & Hedges, 1994). The homogeneity analysis compares the amount of variance in an observed set of effect sizes with the amount of variance that would be expected from sampling error alone (Cooper, Valentine, Charlton, & Melson, 2003). If the total variance (Q_t) is greater than the critical value, considerable variation among the effect sizes exists.

The significance of each factor was tested using two homogeneity tests: a between-group homogeneity test and a within-group homogeneity test. The between-group homogeneity test analyzes the homogeneity of effect sizes across groups. If the between-group variance (Q_b) is greater than the

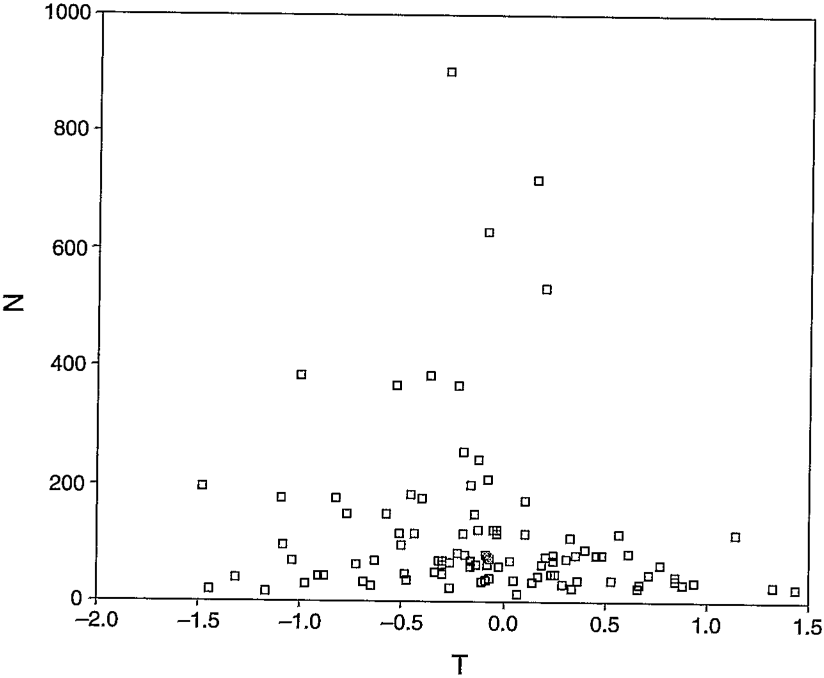

Figure 3. Funnel Plot of the Data Set

critical value, it indicates that there is significant difference among the groups of effect sizes. In other words, the factor is a significant predictor of the effect size difference among different groups. The within-group homogeneity test examines the homogeneity of effect sizes within the groups. If the within-group variance (Q_w) is greater than the critical value, it indicates that the effect sizes within each group are heterogeneous. If the between-group homogeneity test showed the factor was a significant predictor, further analysis was conducted to calculate average effect size and the significant level of each group. The average effect sizes were calculated through univariate analysis of variance using weighted procedures. The effect sizes were weighted by multiplying each independent effect size by the inverse of its variance. The significance level p was calculated through the Z score of the average effect size in each group. We defined a mean effect size as significant when the p value was smaller than .05. The 95% confidence interval of the average effect size also indicated whether it was significant: If the 95% confidence interval included zero, the difference was not significant, and if the 95% confidence interval didn't include zero, the difference was considered significantly positive or negative, depending on the sign of the mean value.

We analyzed the effect sizes using both the fixed-effect model[6] and the random-effect model.[7] But since most of the results from both models are similar, we only present the results from the fixed-effects model in the text and include the results from the random-effects model in Appendix B. The analysis was conducted mainly with Statistical Analysis System (SAS Institute Inc., 1997), and a few descriptive analyses and graphs were generated with the Statistical Package for the Social Sciences (SPSS Inc., 2003).

FINDINGS

In this section we first present our overall finding, followed by results about each of the factors that showed a significant effect on the difference between distance education and face-to-face education.

OVERALL FINDING: IS THERE SIGNIFICANT DIFFERENCE?

The overall weighted mean effect size[8] between distance education and face-to-face education was +0.10, with a 95% confidence interval of [−0.01 0.22] ($z = 1.76$, $p > .05$, $SD = .06$). This finding suggests that when con-

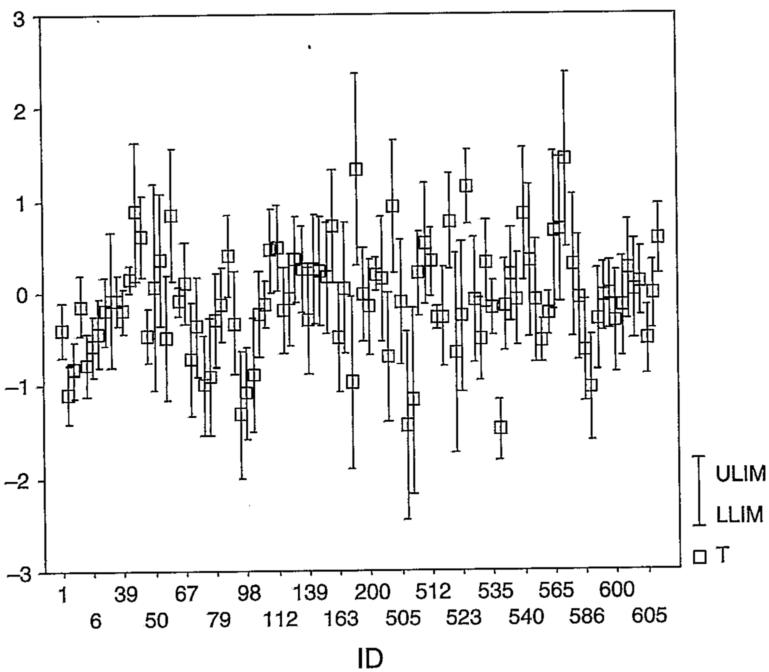

Figure 4. Confidence Interval Plot of the Data Set

sidered as a whole, the studies suggest that there is no significant difference between distance education and face-to-face education, confirming the "no significant difference" claim of previous researchers.

However, a closer look at the data revealed considerable variation among the effect sizes: There is a wide range of effect sizes (from -1.43 to 1.48); about two thirds of the studies show that distance education produced better student outcomes than face-to-face education, whereas the remaining third showed just the opposite. The heterogeneous nature of the effect sizes is clearly shown in the confidence interval plot (Figure 4).[9] A homogeneity test was conducted and the result, $Q(97) = 484.58$, $p < .001$, further confirms the heterogeneous nature of the data set, which means that the effect sizes varied greatly across the studies.

EXPLAINING THE VARIATION: WHAT MAKES DISTANCE EDUCATION EFFECTIVE?

The homogeneity analysis of the data set shows that there is a big variation among the effect sizes. In other words, these studies are very different in terms of their findings. As mentioned previously, the primary goal of this study was to identify factors that may explain why some studies found distance education programs yielded more positive outcomes than their face-to-face counterparts and vice versa. To accomplish this goal, in the following section, we examine how different factors identified in the literature influence the effectiveness of distance education. Table 2 shows all the factors we have identified and the variations they could explain.

In the following section, we present detailed results of the moderator analyses. In the tables included in this section, *k* is the number of inde-

Table 2. Summary of homogeneity tests on each factor

Factor	k	Q_t	Q_b	PQ_b	Q_w	PQ_w
Publication year	97	472.71	76.88	<.0001	395.84	<.0001
Instructor as author	97	472.71	17.12	<.001	455.59	<.0001
Instructor involvement	76	401.98	118.73	<.0001	283.25	<.0001
Student background	78	384.73	26.82	<.0001	357.92	<.0001
Content area	95	459.62	79.08	<.0001	380.54	<.0001
Instructional setting	97	472.71	73.12	<.0001	399.59	<.0001
Interaction type	81	414.03	14.22	<.0001	399.81	<.0001
Media involvement	97	472.71	66.21	<.0001	406.51	<.0001
Measures	97	472.71	43.86	<.0001	428.86	<.0001
Credit	97	472.71	.001	=.9678	472.71	<.0001
Instructor trained	97	472.71	2.32	=.1278	470.39	<.0001
Degree	97	472.71	4.18	=.1236	468.53	<.0001

Note. k = number of effect sizes; $PQ_b = p$ value of Q_b; $PQ_w = p$ value of Q_w.

pendent samples. The *d* statistic is an effect size indicator that represents the standardized mean difference between distance education and face-to-face education. The 95% confidence interval of *d* is also presented together with *d*. The significance level *p* is reported when *d* is significant.

Publication Year

Table 3. Impact of publication year on effectiveness

Publication year	*k*	*d*	*p*
Before 1998	20	$-.10 \pm .11$	
1998 or after	77	$.20 \pm .04$	<.001

As Table 3 suggests, when the study was published has been found to be related to the outcomes of distance and face-to-face education (Machtmes & Asher, 2000). Our analyses confirmed this finding. The year of publication is indeed a significant moderator ($Q_b = 24.40$, $p < .0001$). The year 1998 seems to be an especially important dividing time. Studies published before 1998 did not seem to find a significant difference between distance education and face-to-face education, whereas studies published in and after 1998 found distance education to be significantly more effective than face-to-face education ($d = 0.20$, $p < .001$).

Whether the Instructor Is the Author

As shown in Table 4, whether the author is the instructor of the distance learning course under study makes a significant difference in the effectiveness of distance education compared to face-to-face education ($Q_b = 17.12$, $p < .001$). When the author of the article is also the instructor of the distance education course studied, the outcome seems to significantly favor distance education ($d = 0.33$, $p < .001$); when the author is not the instructor, no significant difference between distance education and face-to-face education is reported; and when the status of author is unknown, distance education is

Table 4. Impact of author status on effectiveness

Instructor is author?	*k*	*d*	*p*
Yes	9	$.33 \pm .15$	<.001
No	18	$-.01 \pm .10$	
Unsure	70	$.18 \pm .12$	<.001

found to be more effective than face-to-face education, but the overall effect size is smaller than in studies whose author is clearly the instructor ($d = 0.18, p < .001$).

Results from the random-effect model suggest a similar situation. The only difference is that when the researchers' status is unknown, studies demonstrate no significant difference between distance education and face-to-face education. The results from both models indicate a possible bias favoring distance education when the author is the instructor of the distance education course.

Outcome Measures

Table 5. Impact of outcome measures on effectiveness

Outcome measures	*k*	*d*	*p*
Grade (including quizzes)	36	0.14 ± .07	<.001
Student attitudes and beliefs	21	0.14 ± .10	<.01
Evaluation of course	20	0.09 ± .11	
Student satisfaction	8	0.14 ± .11	<.01
Student participation	5	0.78 ± .24	<.001
Student evaluation of learning	4	− 0.07 ± .23	
Researcher's observation	2	1.30 ± .71	<.001
Metacognition	1	− 0.11 ± .39	

What is measured or what are used as outcome measures can also have a significant effect on the difference between distance education and face-to-face education (see Table 5). When grades (including quizzes), student attitude and beliefs, student satisfaction, and student participation are measured, or when the outcome is based on researchers' observation, distance learning shows a significantly better outcome than face-to-face learning. When outcome is measured as student evaluation of learning or metacognition, face-to-face education is slightly better than distance learning, but the difference is not significant.

Instructor Involvement

Instructor involvement was the most significant moderator among all the identified factors ($Q_b = 118.726$, $p < .0001$). There were significant differences among studies with different instructor involvement levels. For example, the mean effect sizes of studies at levels 2, 5, 8, and 9 were − 0.07, 0.43 ($p < .001$), 0.47 ($p < .001$), and 0.37 ($p < .001$), respectively. These dif-

Table 6. Impact of instructor involvement level on effectiveness

Groups	*k*	*d*	*p*
Low (≤ 40%)	9	−.24 ± .14	<.001
Medium (50–70%)	32	.29 ± .08	<.001
High (80–100%)	35	.21 ± .06	<.001

ferences suggest a general trend that when instructor involvement is low, the outcomes of distance education are not as positive as those of face-to-face education; when instructor involvement increases, distance education programs yield more positive outcomes than face-to-face education. However, when instructor involvement reaches the highest level, the difference tends to decrease. To further examine this trend, we recoded the variable Instructor Involvement into three levels: low (2, 3, 4), medium (5, 6, 7), and high (8, 9, 10). As shown in Table 6, when instructor involvement is low, face-to-face education is significantly more effective than online education ($d = -0.24$, $p < .001$); when instructor involvement is at the medium level, distance education seems to fare better, and the difference is the largest among these three groups ($d = 0.29$, $p < .001$); and when instructor involvement is high, distance learning still shows a significantly better effect than face-to-face education, but the difference is not as large as at the medium level ($d = 0.21$, $p < .001$).

Student Education Level

Student prior education level is another significant predictor ($Q_b = 26.81$, $p < .001$) of the difference between distance education and face-to-face education. As shown in Table 7, distance education shows significantly better outcomes than face-to-face education with students who have a high school diploma ($d = 0.25$, $p < .001$). But for students with a college degree, the difference between distance education and face-to-face education seems insignificant ($d = 0.06$, $p > .05$).

Table 7. Impact of student prior education level on effectiveness

Student prior education level	*k*	*d*	*p*
Unknown	1	−.20 ± .19	<.05
High school diploma	40	.25 ± .07	<.001
College degree	37	.06 ± .08	

Table 8. Impact of content area on effectiveness

Content area	*k*	*d*	*p*
Business	32	.13 ± .08	<.001
Social science	24	−.11 ± .12	
Science	14	−.03 ± .11	
Computer science	8	.48 ± .11	<.001
Medical science	6	.36 ± .13	<.001
Multiple areas	4	.46 ± .25	<.001
Skills	3	.08 ± .21	
Military	3	.10 ± .23	
Mathematics	1	.09 ± .20	

Content Area

What is being taught is also closely related to whether distance education performs better than its face-to-face counterpart. Our analyses suggest that content area is a significant predictor of the difference between distance education and face-to-face education ($Q_b = 79.08$, $p < .0001$). As shown in Table 8, studies of distance education programs in business, computer science, and medical science found distance learning to be more effective than face-to-face education. In social science and science areas, there is no significant difference between distance learning and face-to-face learning, although face-to-face learning shows a slightly better effect than distance learning. In military, mathematics and specific skills, distance education has a slightly better effect than face-to-face education. Given the small numbers (*k*) of studies in these three areas, it is difficult to draw any firm conclusion. Under the random-effect model, the results demonstrate a similar trend, but the differences in effectiveness between these two learning environments are not significant in almost all of the content areas except in computer science, which shows a significant difference between distance learning and face-to-face learning ($d = 0.50$, $p < .01$).

Level of Instruction

The level of instruction was found to be a significant predictor of the difference between distance learning and traditional learning ($Q_b = 73.12$, $p < .0001$). As shown in Table 9, in multiple settings, face-to-face learning is more effective than distance learning ($d = -0.16$, $p < .01$). However, distance education programs at the undergraduate level and in military settings were more effective than face-to-face learning ($d = 0.36$, $p < .001$; $d = 0.23$, $p < .01$, respectively). There was no significant difference in other

Table 9. Impact of level of instruction on effectiveness

Level	*k*	*d*	*p*
Multiple settings	11	$-.16 \pm .12$	$<.01$
Grades 10–12	1	$.09 \pm .20$	
Associate's degree	6	$.004 \pm .20$	
Undergraduate	36	$.36 \pm .07$	$<.001$
Graduate	35	$.03 \pm .08$	
Professional development	2	$.15 \pm .23$	
Military	6	$.23 \pm .16$	$<.01$

settings. Under the random-effect model, distance learning showed significantly better outcomes than face-to-face learning at undergraduate level ($d = 0.35$, $p < .001$). In all the other settings, there was no significant difference between face-to-face and distance education. The finding that distance programs at the undergraduate level yield more positive learning outcomes than face-to-face programs is closely related to the previous finding that students with high school diplomas seem to benefit more from distance education than students with other prior education levels.

Type of Interaction

The type of interactions, that is, how interactions between the instructor and the students and among students were realized, is a significant predictor of the difference between distance learning and traditional learning ($Q_b = 14.21, p < .0001$). As Table 10 shows, studies of distance programs that employed both synchronous and asynchronous means of interaction found distance education to be significantly better than face-to-face education ($d = .22$, $p < .001$). Although there are two studies that allowed no interaction that found distance education to be more effective than face-to-face education, the small sample size makes the finding questionable. The mean effect size of these two studies is significant under the fixed-effect model ($d = .49$, $p < .01$), but not significant under the random-effect model ($d = .55 \pm .83, p > .05$).

Table 10. Impact of interaction type on effectiveness

Type of interaction	*k*	*d*	*p*
Asynchronous	5	$-.15 \pm .43$	
Synchronous	16	$.08 \pm .10$	
Noninteraction	2	$.49 \pm .40$	$<.01$
Synchronous and asynchronous	57	$.22 \pm .05$	$<.001$

Media Involvement

Table 11. Impact of media involvement level on effectiveness

Media involvement	*k*	*d*	*p*
Medium (6–8)	18	.49 ± .10	<.001
High (9–10)	79	.07 ± .05	<.001

Note. The numbers after "Medium" and "High" represent the media involvement levels. For instance, if the media involvement falls somewhere between 9 and 10, it is considered high media involvement.

The level of media involvement is another significant factor that seems to distinguish the studies in terms of the learning outcomes, $Q_b(96) = 55.07$, $p < .0001$). As shown in Table 11, studies with a coded media involvement of 60–80% reported distance education to be significantly more effective than face-to-face education ($d = 0.50$, $p < .001$). Studies with a 90–100% media involvement also found results favoring distance education, but the difference is much smaller ($d = .07$, $p < .001$). In essence, this finding suggests that distance education mixed with a certain amount of face-to-face instruction seems most effective.

Insignificant Factors

We also examined the effect of other factors but did not find significant effects. These factors include whether the instructor was trained for offering distance education, whether the students took the course for credit, and whether it was a degree course. The results (see Table 2) suggest that these factors did not differentiate studies with results favoring distance education from those favoring face-to-face education.

DISCUSSION

The purpose of this study was not to seek evidence of effectiveness of distance education since, with or without scientific evidence, distance education has already evolved from a marginal form of education to a commonly accepted and increasingly popular alternative to traditional face-to-face education (McIsaac & Gunawardena, 1996; National Center for Education Statistics, 2003; United States General Accounting Office, 2002). Instead, the study was motivated by the pressing need for practical guidance for improving distance education and the dismissive criticism of the immense body of literature in distance education. Rather than waiting for new and improved research and sound theoretical frameworks suggested by some scholars (Gunawardena & McIsaac, 2004; Lockee et al., 1999; McIsaac &

Gunawardena, 1996; Institute for Higher Education Policy, 1999), which have not come out as quickly as expected and may also be deemed outdated and inadequate by some critics as soon as they become available as a result of the rapid changes in distance education technologies, we took a pragmatic approach. We went back to what we already have: the existing literature. We wanted to find out what guidance, if any at all, this literature might have to offer besides the already well-known no-significant-difference conclusion. After searching through thousands of publications, we found a small set of studies that provided sufficient information for secondary analysis. These studies and our analyses seem to support seven findings.

NOT ALL DISTANCE EDUCATION PROGRAMS ARE CREATED EQUAL

Although the aggregated data of all available studies show no significant difference in outcomes between distance education and face-to-face education, there is remarkable difference across the studies. Some studies found distance education to be significantly more effective than face-to-face, whereas others found the opposite to be true. This finding is consistent with previous research and supports the popular impression of distance education, in that distance education as a form of education is as good (or as bad) as face-to-face education. It highlights, however, an important and often neglected fact about distance education literature: Distance education programs, just like traditional education programs, vary a great deal in their outcomes. Thus it is advisable not to automatically apply the "no-significant-difference" label to all distance education programs just because the positive findings of some studies cancel out the negative findings of other studies.

INTERACTION IS KEY TO EFFECTIVE DISTANCE EDUCATION

Whether and how much students interact with peers and instructors seems to be a differentiating quality of distance programs in terms of learning outcomes. Our analysis found three interaction-related factors that tend to distinguish studies with outcomes favoring distance education from those favoring face-to-face education: instructor involvement, media involvement, and types of interactions. The extent to which the instructor is involved in the actual delivery of the content directly influences opportunities for students to interact with the instructor. The study found that studies with higher instructor involvement tend to favor distance education over its face-to-face counterpart. Whether a distance course included a face-to-face component, which presumably increased the opportunity for interaction, was also found to set studies favoring distance education apart from those favoring face-to-face education. Perhaps the most compelling and direct evidence is the distinguishing power of the types of interactions. The study clearly shows that studies of distance programs with both synchronous and asynchronous interactions reported more positive outcomes than those with only one type of interaction.

Interaction among students and between instructors and students is the hallmark of education. But distance education has traditionally been at a disadvantage in this aspect. Either because of the limitation of technology or because of cost, distance education programs, until recently, have not been able to offer the full range of communication channels to students and instructors. With the advent of more cost-effective and efficient communication technologies such as the Internet, distance education programs have started to provide both synchronous and asynchronous communications that enable a broad range of interactions between students and instructors and among students. Recent distance education programs that took advantage of these tools have reported positive outcomes (for example, see Institute for Higher Education Policy, 2000; Levin, Levin, & Chandler, 2001; Twigg, 2001).

There are, however, nontechnological costs associated with offering both synchronous and asynchronous interactions. The first is that someone needs to coordinate and manage the interactions. It is difficult to imagine that the provision of technological capacity for interaction alone automatically leads to meaningful interactions. The instructor needs to be present to answer student questions and facilitate discussions. Second, someone needs to maintain the infrastructure for both synchronous and asynchronous communications. Although there are plenty of communication tools available, they need to be maintained and updated. Third, both students and instructors need training and help with these tools. In traditional distance education, delivery is the main concern. The instructor and students may not need to know how to actually operate the tools because there is a technician to handle the transmission and reception of content. In interactive distance education, both instructors and students need to use the communication tools, which are often computer software unfamiliar to instructors and students. Thus training and support become necessary.

LIVE HUMAN INSTRUCTORS ARE NEEDED IN DISTANCE EDUCATION

Distance education comes in all forms and shapes. Some programs are Web-based, computerized instructional courses that do not have any involvement of a "live" instructor. Students simply interact with the computer. Some programs take the form of broadcasting prerecorded videos of instruction, and these do not have instructor involvement either. Still there are programs that take the shape of correspondence courses, but with limited e-mail communication with a "live" instructor, who may have to supervise hundreds of students. There are also programs that are just like traditional face-to-face education, in that there are scheduled class meetings and out-of-class office hours, except that these meetings take place in an online environment. Our findings suggest that the presence of a "live" instructor is important for effective distance education. We found that the degree of instructor involvement is a significant distinguishing quality of effective and ineffective distance education programs. Although this finding may be disappointing to those who think distance education may be more efficient, in terms of personnel cost, we want to emphasize that, based on previous research, "live" human instructors are still needed to ensure quality distance education.

THE RIGHT MIXTURE OF HUMAN AND TECHNOLOGY SEEMS MOST BENEFICIAL

Today's distance education inevitably involves the use of some kind of technology, be it interactive television or the Internet. In fact one of the defining characteristics of distance education is its use of technology to remove the distance between the provider and recipient of instruction. Our analysis suggests that those studies that used a combination of technology and face-to-face education resulted in the most positive outcomes.

Recent research supports a hybrid model of distance education that combines both a face-to-face component and a technology-mediated distance component. For instance, in reviewing and projecting the use and influence of information technology on learning and teaching, Lant (2002) expressed the necessity of treating online technologies and traditional learning as complementary to each other. Although it may not be possible for all distance programs to include a face-to-face component, there are tools (e.g., video conferencing) that can effectively remove the distance and create effective social organizations (Levin et al., 2001).

DISTANCE EDUCATION MAY BE MORE APPROPRIATE FOR CERTAIN CONTENT

The nature of what is being taught, that is, the content area or the curriculum, seems to affect the effectiveness of distance education as well. Studies of distance courses in computer science, for example, reported more positive results than other content areas. It was also found that studies of college-level courses had results that favor distance education over face-to-face, while studies of graduate level programs found less positive results. Typically college level courses differ from graduate level courses in terms of content and desirable outcomes. This suggests, although more studies are needed to confirm, that distance education may be more effective in teaching some content than others.

Another related finding is that studies of programs whose students had high school diplomas found distance education to be significantly more effective than face-to-face education, while studies of other students did not find such strong effect. One possible explanation of this finding has to do the content of the courses. We can assume that most of the courses that the students with a high school diploma take are college level courses, and those taken by students with a college degree are graduate level courses. Relatively speaking, college level courses could have more of a focus on knowledge and skill acquisition, while graduate level courses focus more on idea or research interest development. It is possible that knowledge and skills can be taught more effectively in distance education, but the development of an idea or research interest may need more discussion and interactions with the instructor and other students. In other words, the advantage of distance education in delivering learning content in college level courses may not hold for graduate level courses where more complex ideas are explored.

SOME LEARNERS MAY BE MORE ABLE TO TAKE ADVANTAGE OF DISTANCE EDUCATION

The finding that some learners may be more able to take advantage of distance education than others is not new. A number of studies (e.g., Institute for Higher Education Policy, 1999) have suggested that students with certain qualities seem to benefit more from distance education. In the studies we analyzed, there was not sufficient information about learner characteristics to examine how individual characteristics affect learning outcomes of distance education. We were only able to conduct analysis of prior education level, which may interact with many other learner characteristics, such as gender, study habits, learning styles, learning environment, access to resources, experiences with distance learning, and technology proficiency. Thus we are not able to draw any conclusion. However, the finding that a high school diploma seems to distinguish effective distance education programs from ineffective ones leads to the suggestion that more attention should be paid to learners in future studies.

DISTANCE EDUCATION SEEMS TO GET BETTER

The finding that studies prior to 1998 found distance education to be less effective than face-to-face education, whereas those post-1998 found the opposite, could be an indication that distance programs are getting better—with more powerful delivery media and more sophisticated support systems. There were some significant changes in technology employed in distance education in the mid-1990s. The 1997–98 report produced by National Center for Education Statistics (1999) reported that among all higher education institutions offering any distance education, the percentages of institutions using two-way interactive video and one-way prerecorded video were essentially the same in 1997–98 as in 1995. However, with the dramatic growth of the World Wide Web, technologies based on the Internet, such as e-mail, Web page, and Web board have increased rapidly since 1998. The change on technology employed in distance learning is likely to influence many aspects of distance education such as how learning materials are presented and how the teacher and students communicate and interact and hence to influence the effectiveness of distance education.

The difference could also be a result of the maturation of distance education programs. Over the years, distance education programs have become more mature, with an increased amount and variety of support for both instructors and students. For example, distance education programs are providing technical help through a variety of means, including toll-free phones, e-mail, real-time chat rooms, and online tutorials for technical assistance (Institute for Higher Education Policy, 2000; Levin et al., 2001). Meanwhile, the instructors are receiving more training and becoming more experienced with teaching online courses. Moreover, students are becoming more comfortable working with computers and learning online. All these factors can contribute to the increased effectiveness on distance education in recent years.

Another possible, but less likely, explanation is that there was a paradigm shift: Distance education has been accepted as an effective form of education, and thus only studies that report positive findings have been published.

CONCLUSION

There are frequent calls for new conceptual and theoretical framework for distance education research and practice (Gunawardena & McIsaac, 2004; Head, Lockee, & Oliver, 2002; Lock, 2002; McIsaac & Gunawardena, 1996). Although such exercises may be useful, they may be unnecessary. Distance education is in essence still education (Shale, 1990). Results from this study further support this argument. The factors found to have an impact on the effectiveness of distance education are also factors that would affect the effectiveness of face-to-face education. Additionally, the one factor that often sets distance education apart from face-to-face education—distance or the technology that is used to remove the effect of distance—is quickly disappearing as face-to-face distance education increasingly uses technology to support teaching and learning. In other words, the line between distance education and face-to-face education is quickly being erased (McIsaac & Gunawardena, 1996).

When distance education is considered the same as face-to-face education, we are encouraged to consider the abundance of theoretical, analytical, and conceptual frameworks for understanding education. Schwab's four common places—teacher, student, what is taught, and milieux of teaching-learning—can serve as a very useful overarching framework for studying and thinking about distance education, as suggested by the present study. There are other more detailed and specific theoretical frameworks that can help us understand the relationships and interactions among the variables located in the four common places.

Although it is desirable to have well-designed true random experimental and longitudinal studies that ask the right questions in distance education research, as expected by some scholars, such high-quality studies are hard to come by for a number of practical reasons (Cook & Campbell, 1979). This is the reality of research in social sciences in general and distance education research in particular. Thus we have to accept this reality and find novel approaches to understand existing research. The study reported here is an example of how to take a pragmatic approach to research synthesis in order to answer pressing and practical questions.

Finally, this study is by no measure conclusive because of a number of limitations. For example, the exclusion of publications from other sources, such as dissertation and conference proceedings, may have biased the findings. One of the primary purposes of this study is to test a new way of working with previous research. Although we believe the findings suggest that examining factors that moderate the effects of distance education can lead to the identification of practical guidance for practice and research, we also encourage future studies to include a more complete examination of all studies and integrate analysis of qualitative studies from the literature.

APPENDIX A

LIST OF STUDIES

Table A1. List of studies and their features

Author	Year	Indicator	Instructor involvement[a]	Involvement type[b]	Media involvement[a]	Content area	Learner background	Author is instructor	Sample size	t^c
Arbaugh, J. B.	2000	Grades	8	4	10	Business	College degree	Not sure	60	1.33
Arbaugh, J. B.	2000	Student attitude and beliefs	8	4	10	Business	College degree	Not sure	60	−0.02
Bader, Mary B.; Roy, Sam	1999	Student evaluation of course	6	4	10	Business	College degree	Not sure	24	0.34
Barkhi, Reza; Brozovsky, John	2000	Grades	6	4	10	Business	College degree	Not sure	62	0.77
Barkhi, Reza; Brozovsky, John	2000	Student evaluation of course on delivery	6	4	10	Business	College degree	Not sure	63	0.19
Barkhi, Reza; Brozovsky, John	2000	Student evaluation of course on resolving conflicts	6	4	10	Business	College degree	Not sure	64	−0.08
Barkhi, Reza; Brozovsky, John	2000	Student participation	6	4	10	Business	College degree	Not sure	64	−0.14
Boulet, Marie-Michele; Boudreault, Serge; Guerett	1998	Fundamental grades	5	4	8	Computer science	High school diploma	Not sure	176	−0.4
Boulet, Marie-Michele; Boudreault, Serge; Guerett, Louis	1998	Fundamental grades	8	4	6	Computer science	High school diploma	Not sure	150	−0.14
Boulet, Marie-Michele; Boudreault,	1998	Grades	8	4	6	Computer science	High school diploma	Not sure	150	−0.58

Table A1. (Continued)

Author	Year	Indicator	Instructor involvement[a]	Involvement type[b]	Media involvement[a]	Content area	Learner background	Author is instructor	Sample size	t^c
Serge; Guerett, Louis										
Boulet, Marie-Michele; Boudreault, Serge; Guerett, Louis	1998	Problem solving Grades	5	4	8	Computer science	High school diploma	Not sure	176	−0.83
Boulet, Marie-Michele; Boudreault, Serge; Guerett, Louis	1998	Student attitude and beliefs	5	4	8	Computer science	High school diploma	Not sure	176	−1.1
Boulet, Marie-Michele; Boudreault, Serge; Guerett, Louis	1998	Student attitude and beliefs	8	4	6	Computer science	High school diploma	Not sure	150	−0.78
Braun, Kathryn L.; Roberts, Ellen; Dubanoski, Joan	1998	Grades	6	4	8	Social science		Not sure	68	−0.32
Braun, Kathryn L.; Roberts, Ellen; Dubanoski, Joan	1998	Student attitude and beliefs	6	4	8	Social science		Not sure	68	−0.16
Braun, Kathryn L.; Roberts, Ellen; Dubanoski, Joan	1998	Student attitude and beliefs	6	4	8	Social science		Not sure	68	0.25
Braun, Kathryn L.; Roberts, Ellen; Dubanoski, Joan	1998	Student attitude and beliefs	6	4	8	Social science		Not sure	68	0.03
Braun, Kathryn L.; Roberts, Ellen; Dubanoski, Joan	1998	Student evaluation of course on reading	6	4	8	Social science		Not sure	68	−1.04
Braun, Kathryn L.; Roberts, Ellen; Dubanoski, Joan	1998	Student evaluation of course on videos	6	4	8	Social science		Not sure	68	−0.63

Braun, Kathryn L.; Roberts, Ellen; Dubanoski, Joan	1998	Student satisfaction	6	4	8	Social science		Not sure	68	−0.3
Cheng, Hui-Chuan, et al.	1991	Grades	5	4	10	Social science	College degree	Not sure	35	0.36
Cheng, Hui-Chuan, et al.	1991	Grades		4	10	Social science	College degree	Not sure	35	−0.48
Cheng, Hui-Chuan, et al.	1991	Student attitude and beliefs	5	4	10	Social science	College degree	Not sure	35	0.84
Cheng, Hui-Chuan, et al.	1991	Student attitude and beliefs		4	10	Social science	College degree	Not sure	27	−0.97
Cheng, Hui-Chuan, et al.	1991	Student participation	5	4	10	Social science	College degree	Not sure	37	−0.48
Cheng, Hui-Chuan, et al.	1991	Student participation		4	10	Social science	College degree	Not sure	29	0.05
Clark, Ruth Anne; Jones, David	2001	Student evaluation of course	5	4	9	Skills	High school diploma	No	123	−0.05
Clark, Ruth Anne; Jones, David	2001	Student evaluation of learning	5	4	9	Skills	High school diploma	No	123	−0.03
Collins, Michael	2000	Grades	7	4	10	Military	High school diploma	Yes	256	−0.2
Collins, Michael	2000	Grades	7	4	10	Military	High school diploma	Yes	173	0.11
Dellana, Scott A.; Collins, William H.; West, David	2000	Grades	8	4	10	Business	High school diploma	Not sure	199	−0.16
Dellana, Scott A.; Collins, William H.; West, David	2000	Student participation	8	4	10	Business	High school diploma	Not sure	196	−1.48
Furst-Bowe, Julie A.	1997	Student evaluation of course	9	4	10	Business	College degree	Yes	40	−0.07
Hacker, Roger; Sova, Brian	1998	Grades	10	3	10	Social science	College degree	Not sure	43	−0.91
Holman, Lucy	2000	Quizzes	3	3	10	Skills	High school diploma	No	83	−0.23
Johnson, Margaret	2002	Grades	10	4	10	Science	High school diploma	Not sure	116	−0.44

Table A1. (Continued)

Author	Year	Indicator	Instructor involvement[a]	Involvement type[b]	Media involvement[a]	Content area	Learner background	Author is instructor	Sample size	t^c
Johnson, Margaret	2002	Metacognition	10	4	10	Science	High school diploma	Not sure	116	0.11
Johnson, Margaret	2002	Student attitude and beliefs on confidence	10	4	10	Science	High school diploma	Not sure	116	−0.03
Johnson, Margaret	2002	Student attitude and beliefs on liking group work	10	4	10	Science	High school diploma	Not sure	116	−0.52
Johnson, Margaret	2002	Student attitude and beliefs on liking using computers	10	4	10	Science	High school diploma	Not sure	116	0.57
Johnson, Margaret	2002	Student satisfaction	10	4	10	Science	High school diploma	Not sure	116	−0.2
King, Peter; Hi-ldreth, David	2001	Grades	7	4	10	Science	High school diploma	Not sure	76	−0.08
Kirman, J. M.; Goldberg, J.	1982	Grades	10	4	10	Science		No	718	0.15
Kirman, J. M.; Goldberg, J.	1982	Student participation	10	4	10	Science		No	30	0.89
Leasure, A. Renee; Davis, Lisa; Thie-von, Susan L.	2000	Grades	3	4	8	Medical science	High school diploma	Not sure	66	−0.27
Leasure, A. Renee; Davis, Lisa; Thie-von, Susan L.	2001	Student attitude and beliefs	3	4	7	Medical science	High school diploma	No	587	−0.28
Lia-Hoagberg, Betty; Vellenga, Barbara; Miller, M	1999	Student evaluation of course	9	2	10	Medical science	College degree	Not sure	369	−0.23
Lia-Hoagberg, Betty; Vellenga, Barbara; Miller, Marilee	1999	Student satisfaction	9	2	10	Medical science	College degree	Not sure	369	−0.53

Magiera, Frank T.	1994	Grades	8	4	9	Business	High school diploma	Not sure	35	−0.09
McGreal, Rory	1994	Student attitude and beliefs	10	2	10			Not sure	62	−0.3
Miller, John W., et al.	1993	Grades	10	2	10	Social science	College degree	Not sure	22	0.66
Miller, John W., et al.	1993	Grades	10	2	10	Social science	College degree	Not sure	29	0.67
Miller, John W., et al.	1993	Student evaluation of course	10	2	10	Social science	College degree	Not sure	22	1.43
Miller, John W., et al.	1993	Student evaluation of course	10	2	10	Social science	College degree	Not sure	29	0.29
Murphy, Tim H.	2000	Grades	7	4	10	Science	High school diploma	No	80	−0.09
Nesler, Mitchell S.; Hanner, Mary Beth; Melburg, Valerie	2001	Student attitude and beliefs	3	4	7	Social science	High school diploma	Not sure	905	0.53
Nesler, Mitchell S.; Hanner, Mary Beth; Melburg, Valerie	2001	Student attitude and beliefs	3	4	7	Social science	High school diploma	Not sure	318	0.33
Phelps, Ruth H., et al.	1991	Grades	10	4	10	Engineering	High school diploma	No	384	−0.37
Phelps, Ruth H., et al.	1991	Grades	10	4	10	Skills	High school diploma	No	384	−1.00
Phelps, Ruth H., et al.	1991	Student evaluation of learning	10	4	10	Engineering	High school diploma	No	63	−0.72
Phillips, Melodie R.; Peters, Mary Jane	1999	Student evaluation of course	8	4	7	Business	High school diploma	Not sure	95	−0.51
Redding, Terrence R.; Rotzien, Jack	2001	Grades	5	4	10	Business		Yes	40	−1.31
Redding, Terrence R.; Rotzien, Jack	2001	Grades	5	4	10	Business		Yes	96	−1.08
Redding, Terrence R.; Rotzien, Jack	2001	Grades	5	4	10	Business		Yes	44	−0.88

Table A1. (Continued)

Author	Year	Indicator	Instructor involvement[a]	Involvement type[b]	Media involvement[a]	Content area	Learner background	Author is instructor	Sample size	t^c
Roblyer, M. D.	1999	Student evaluation of course on control of learning pace	2	1	10	Science	High school diploma	No	33	0.15
Roblyer, M. D.	1999	Student evaluation of course on interaction	2	1	10	Science	High school diploma	No	33	−0.69
Roblyer, M. D.	1999	Student evaluation of course on logistic issues	2	1	10	Science	High school diploma	No	33	0.2
Roblyer, M. D.	1999	Student evaluation of course on technology	2	1	10	Science	High school diploma	No	33	0.94
Ryan, Walter F.	1996	Grades	10	4	10	Mathematics		Not sure	630	−0.09
Schoenfeld-Tacher, Regina; McConnell, Sherry; Gra, Michelle	2001	External evaluation on student interactions	6	2	10	Medical science	High school diploma	No	18	−1.44
Schoenfeld-Tacher, Regina; McConnell, Sherry; Gra, Michelle	2001	External evaluation on teacher–student interaction	6	2	10	Medical science	High school diploma	No	19	−0.11
Schoenfeld-Tacher, Regina; McConnell, Sherry; Gra, Michelle	2001	Grades	6	2	10	Medical science	High school diploma	No	42	0.84
Simpson, Henry, et al.	1991	Grades	10	4	10	Military		Not sure	13	0.07
Smeaton, Alan F.; Keogh, Gary	1999	Grades	5	4	10	Computer science	High school diploma	Yes	243	−0.13
Sonner, Brenda S.	1999	Grades	4	1	8	Business	High school diploma	Not sure	72	0.31
Sowinski, Kevin M.; Scott, Steven	2000	Student attitude and beliefs	5	0	10	Computer science	College degree	Not sure	122	−0.13

A.; Carlstedt, Bruce									
Spooner, Fred; Jordan, Luann; Al-gozzine, Bob; Spooner, Melba	1999	Student evaluation of course	2	8	Social science	College degree	Yes	27	−0.65
Spooner, Fred; Jordan, Luann; Al-gozzine, Bob; Spooner, Melba	1999	Student evaluation of course	2	8	Social science	College degree	Yes	24	−0.26
Sugrue, Brenda; Rietz, Thomas; Hansen, Sarah	1999	Grades		10	Business	College degree	Not sure	90	0.4
Sugrue, Brenda; Rietz, Thomas; Hansen, Sarah	1999	Grades		10	Business	College degree	Not sure	51	−0.33
Sugrue, Brenda; Rietz, Thomas; Hansen, Sarah	1999	Student attitude and beliefs		10	Business	College degree	Not sure	72	−0.19
Sugrue, Brenda; Rietz, Thomas; Hansen, Sarah	1999	Student attitude and beliefs		10	Business	College degree	Not sure	80	0.36
Sugrue, Brenda; Rietz, Thomas; Hansen, Sarah	1999	Student attitude and beliefs on anticipated grade		10	Business	College degree	Not sure	47	0.72
Sugrue, Brenda; Rietz, Thomas; Hansen, Sarah	1999	Student attitude and beliefs on difficulty		10	Business	College degree	Not sure	80	0.49
Sugrue, Brenda; Rietz, Thomas; Hansen, Sarah	1999	Student attitude and beliefs on difficulty		10	Business	College degree	Not sure	46	0.26
Sugrue, Brenda; Rietz, Thomas; Hansen, Sarah	1999	Student attitude and beliefs on effort		10	Business	College degree	Not sure	43	0.24
Sugrue, Brenda; Rietz, Thomas; Hansen, Sarah	1999	Student attitude and beliefs on perceived value		10	Business	College degree	Not sure	81	0.46

Table A1. (Continued)

Author	Year	Indicator	Instructor involvement[a]	Involvement type[b]	Media involvement[a]	Content area	Learner background	Author is instructor	Sample size	t^c
Sugrue, Brenda; Rietz, Thomas; Hansen, Sarah	1999	Student attitude and beliefs on perceived value			10	Business	College degree	Not sure	47	−0.3
Sugrue, Brenda; Rietz, Thomas; Hansen, Sarah	1999	Student evaluation of learning			10	Business	College degree	Not sure	80	−0.07
Sugrue, Brenda; Rietz, Thomas; Hansen, Sarah	1999	Student evaluation of learning			10	Business	College degree	Not sure	45	0.17
Sugrue, Brenda; Rietz, Thomas; Hansen, Sarah	1999	Student satisfaction			10	Business	College degree	Not sure	47	0.25
Summers, Marcia; Anderson, Jennie L.; Hines, Allyn R.; Gelder, Barbara C.; Dean, Raymond S.	1996	Student satisfaction	4	2	10			Not sure	532	−0.16
Thyer, Bruce A.; Artelt, Thomas; Markward, Martha	1998	Student evaluation of course	10	2	10	Social science	High school diploma	Not sure	76	−1.17
Thyer, Bruce A.; Artelt, Thomas; Markward, Martha	1998	Student evaluation of course	10	2	10	Social science	High school diploma	Not sure	110	0.21
Wang, Alvin Y.; Newlin, Michael H.	2000	Grades	4	4	9	Social science	High school diploma	Not sure	115	1.14
Wilkinson, Kelly L.; Hemby, K. Virginia	2000	Student satisfaction			10	Business		Not sure	84	0.61

Wilkinson, Kelly L.; Hemby, K. Virginia	2000	Student satisfaction			10	Business	Not sure	183	−0.46
Wisher, Robert A.; Curnow, Christina K.	1999	Grades	10	4	10	Computer science	No	208	−0.08

[a]Both instructor involvement and media involvement were rated on a scale from 1 to 10.

[b]1 = asynchronous interactive involvement; 2 = synchronous interactive involvement; 3 = noninteractive involvement; 4 = synchronous interactive involvement and asynchronous interactive involvement.

[c]Weighted effect size.

APPENDIX B

RESULTS UNDER THE RANDOM-EFFECT MODEL

Table B1. Impact of Publication Year on Effectiveness

Group	*k*	*d*	*p*
Before 1998	20	− .10 ± .27	
1998 or after	77	.15 ± .12	<.01

Table B2. Impact of author status on effectiveness

Author status	*k*	*d*	*p*
Yes	9	.47 ± .38	<.01
No	18	.13 ± .26	
Unsure	70	.05 ± .13	

Table B3. Impact of outcome measures on effectiveness

Outcome measures	*k*	*d*	*p*
Grade (including quizzes)	36	0.14 ± .18	
Student attitude and beliefs	21	0.04 ± .24	
Evaluation of course	20	0.02 ± .26	
Student satisfaction	8	0.07 ± .37	
Student participation	5	0.49 ± .52	
Student evaluation of learning	4	− 0.08 ± .55	
Researcher's observation	2	1.30 ± .99	<.05
Metacognition	1	− 0.11 ± 1.06	

Table B4. Impact of instructor involvement on effectiveness

Groups	*k*	*d*	*p*
Low (≤ 40%)	9	− .12 ± .36	
Medium (50–70%)	32	.25 ± .20	<.01
High (80–100%)	35	.14 ± .20	

Table B5. Impact of learner background on effectiveness

Learner background	*k*	*d*	*p*
Unknown	1	.20 ± .98	
High school diploma	40	.21 ± .17	<.01
College degree	37	−.07 ± .19	

Table B6. Impact of content area on effectiveness

Content area	*k*	*d*	*p*
Business	32	.07 ± .20	
Social science	24	−.09 ± .25	
Science	14	−.03 ± .30	
Computer science	8	.50 ± .37	<.01
Medical science	6	.36 ± .43	
Multiple areas	4	.54 ± .56	
Skills	3	.10 ± .62	
Military	3	.03 ± .67	
Mathematics	1	.09 ± 1.01	

Table B7. Impact of level of instruction on effectiveness

Level	*k*	*d*	*p*
Multiple settings	11	−.18 ± .36	
Grade 10–12	1	.09 ± 1.0	
Associate's degree	6	−.03 ± .46	
Undergraduate	36	.35 ± .18	<.001
Graduate	35	−.11 ± .19	
Professional development	2	−.05 ± .74	
Military	6	.35 ± .44	

Table B8. Impact of interaction type on effectiveness

Interaction type	*k*	*d*	*p*
Asynchronous	5	−.13 ± .55	
Synchronous	16	−.05 ± .31	
Noninteraction	2	.55 ± .83	
Synchronous and asynchronous	57	.21 ± .15	<.01

Table B9. Impact of media involvement level on effectiveness

Media involvement	k	d	p
Medium	18	$.41 \pm .25$	$<.001$
High	79	$.03 \pm .13$	

The study was conducted as a project of the Center for Teaching and Technology in the College of Education at Michigan State University. The views expressed in the article do not necessarily represent those of the college or the university. We would like to thank Tom Bird, Richard Banghart, Mark Urban-Lurain, and Gary Cziko for sharing their insights on the issues of distance education and educational technology during formal and informal discussions. Ou (Lydia) Liu, Yun Fan, and Eduardo Junqueira Rodrigues assisted with data collection and coding. We would particularly like to thank Meng-jia Wu and Ken Frank for their help with the meta-analysis.

Notes

1 In this article, *distance education* is used to refer to both traditional distance education and more recent forms of distance education, which are more commonly referred to as *online education, online learning, Web-based learning,* or *Web-based education.*

2 $d = M_1 - M_2/\sigma_{\text{pooled}}$, where $\sigma_{\text{pooled}} = \sqrt{[(\sigma_1^2 + \sigma_2^2)/2]}$, M_1 is the mean of experimental group, and M_2 is the mean of control group.

3 $d = 2t/\sqrt{(df)}$.

4 $g = \sqrt{F * (\frac{1}{N_e} + \frac{1}{N_c})}$, where N_e is the sample size of the experimental group and N_c is the sample size of the control group.

5 $J = 1 - (3/(4 * (N_e + N_c - 2) - 1))$ and $T = g * J$, where g is the initial effect size computed from primary studies, N_e is the sample size of the experimental group, N_c is the sample size of the control group, and T is the unbiased effect size.

6 Under the fixed-effect model, all variation is conceived as being due to sampling error, and we are estimating a "common" effect.

7 Under the random-effect model, we assume studies are from different populations. Each population may have a different effect, and we estimate the amount of uncertainty due to those differences, and thus we are estimating an "average" effect.

8 Since the overall homogeneity test indicated that the effect sizes were from different populations, the random-effect model was applied to calculate the weighted overall mean effect size.

9 The confidence interval plot shows the 95% confidence interval of each effect size. If the effect sizes are homogeneous, the confidence intervals will be distributed evenly, without much variation, and a line—the mean—can be drawn through all the confidence intervals.

References

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.

*Arbaugh, J. B. (2000). Virtual classroom versus physical classroom: An exploratory study of class discussion patterns and student learning in an asynchronous Internet-based MBA course. *Journal of Management Education, 24*(2), 213–233.

*Bader, M. B., & Roy, S. (1999). Using technology to enhance relationships in interactive television classrooms. *Journal of Education for Business, 74*(6), 357–362.

*Barkhi, R., & Brozovsky, J. (2000). An analysis of the dynamics of a distance course. *Journal of Educational Technology Systems, 28*(4), 311–325.

Bates, A. W. (1995). *Technology, open learning and distance education.* London: Routledge.

Becker, B. J. (2000). Multivariate meta-analysis. In H. A. Tinsley & S. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 499–525). Burlington, MA: Academic Press.

Begg, C. B. (1994). Publication bias. In *The handbook of research synthesis* (pp. 399–409). New York: Russell Sage Foundation.

*Boulet, M.-M., Boudreault, S., & Guerett, L. (1998). Effects of a television distance education course in computer science. *British Journal of Educational Technology, 29*(2), 101–111.

*Braun, K. L., Roberts, E., & Dubanoski, J. (1998). Evaluating a telecourse on aging. *Educational Gerontology, 24*(2), 141–157.

*Carl, D. R., & Densmore, B. (1988). Introductory accounting on Distance University Education via Television (DUET): A Comparative Evaluation. *Canadian Journal of Educational Communication, 17*(2), 81–94.

*Carr-Chellman, A. A., Dyer, D., & Breman, J. (2000). Burrowing through the network wires: Does distance detract from collaborative authentic learning? *Journal of Distance Education, 15*(1), 39–62.

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

*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.

*Chisholm, M. A., Miller, A. W., & Spruill, W. J. (2000). Influence of interactive videoconferencing on the performance of pharmacy students and instructors. *American Journal of Pharmaceutical Education, 64*(2), 152–158.

*Clark, R. A., & Jones, D. (2001). A comparison of traditional and online formats in a public speaking course. *Communication Education, 50*(2), 109–124.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445–459.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research & Development, 42*(2), 21–29.

Clark, R. E., & Salomon, G. (1986). Media in teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 464–478). New York: MacMillan.

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

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings.* Boston: Houghton Mifflin.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews.* Thousand Oaks, CA: Sage.

Cooper, H., Valentine, J. C., Charlton, K., & Melson, A. (2003). The effects of modified school calendars on student achievement and on school and community attitudes. *Review of Educational Research, 73*(1), 1–52.

*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 Agricommunication course. *Journal of Agricultural Education, 39*(4), 65–75.

*Dellana, S. A., Collins, W. H., & West, D. (2000). On-line education in a management science course—Effectiveness and performance factors. *Journal of Education for Business, 76*(1), 43–47.

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association, 263*(10), 1385–1389.

Dubin, R., & Hedley, R. (1969). *The medium may be related to the message.* Eugene: University of Oregon Press.

*Frear, V., & Hirschbuhl, J. J. (1999). Does interactive multimedia promote achievement and higher level thinking skills for today's science students? *British Journal of Educational Technology, 30*(4), 323–329.

*Furst-Bowe, J. A. (1997). Comparison of student reactions in traditional and videoconferencing courses in training and development. *International Journal of Instructional Media, 24*(3), 197–205.

Glaser, B. G. (1992). *Basics of grounded theory analysis: Emergence vs. forcing.* Mill Valley, CA: Sociology Press.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Chicago: Aldine.

Glass, V. G. (1976). Primary, secondary, and meta-analysis research. *Educational Researcher, 5*(10), 3–8.

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

Gunawardena, C. N., & McIsaac, M. S. (2004). Distance education. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (2nd ed., pp. 355–396). Mahwah, NJ: Erlbaum.

*Hacker, R., & Sova, B. (1998). Initial teacher education: A study of the efficacy of computer mediated courseware delivery in a partnership context. *British Journal of Educational Technology, 29*(4), 333–341.

Head, J. T., Lockee, B. B., & Oliver, K. M. (2002). Method, media, and mode: Clarifying the discussion of distance education effectiveness. *Quarterly Review of Distance Education, 3*(3), 261–268.

*Holman, L. (2000). A comparison of computer-assisted instruction and classroom bibliographic instruction. *Reference & User Services Quarterly, 40*(1), 53–60.

Institute for Higher Education Policy. (1999). *What's the difference? A review of contemporary research on the effectiveness of distance learning in higher education.* Washington DC: Author.

Institute for Higher Education Policy. (2000). *Quality online: Benchmarks for success in Internet-based distance education.* Washington, DC: National Education Association.

ISI Research Soft. (2001). EndNote (version 5.0) [Computer software]. Berkeley, CA: Author.

*Jackman, D. H., & Swan, M. K. (2000). Comparing the success of students enrolled in distance education courses vs. face-to-face classrooms. *Journal of Technology Studies, 26*(1), 58–63.

*Johnson, M. (2002). Introductory biology "online": Assessing outcomes of two student populations. *Journal of College Science Teaching, 31*(5), 312–317.

Joy, E. H., & Garcia, F. E. (2000). Measuring learning effectiveness: A new look at no-significant-difference findings. *Journal of Asynchronous Learning Networks, 4*(1), 33–39.

*King, P., & Hildreth, D. (2001). Internet courses: Are they worth the effort? *Journal of College Science Teaching, 31*(2), 112–115.

*Kirman, J. M., & Goldberg, J. (1982). Distance education: Simultaneous one-way television and group telephone conferencing for satellite map instruction. *Alberta Journal of Educational Research, 28*(1), 51–57.

Koumi, J. (1994). Media comparison and deployment: A practitioner's view. *British Journal of Educational Technology, 25*, 41–57.

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

*Leasure, A. R., Davis, L., & Thievon, S. L. (2000). Comparison of student outcomes and preferences in a traditional vs. World Wide Web–based baccalaureate nursing research course. *Journal of Nursing Education, 39*(4), 149–154.

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

*Lia-Hoagberg, B., Vellenga, B., & Miller, M. (1999). A partnership model of distance education: Students' perceptions of connectedness and professionalization. *Journal of Professional Nursing, 15*(2), 116–122.

Lipsey, M. W. (1994). Identifying potentially interesting variables and analysis opportunities. In H. Cooper and L. V. Hedges (Eds.). *The handbook of research synthesis* (pp. 111–123). New York: Russell Sage Foundation.

Lock, J. V. (2002). Laying the Groundwork for the Development of Learning Communities within Online Courses. *Quarterly Review of Distance Education, 3*(4), 395–408.

Lockee, B. B., Burton, J. K., & Cross, L. H. (1999). No comparison: Distance education finds a new use for "no significant difference." *Educational Technology Research and Development, 7*(3), 33–42.

Machtmes, K., & Asher, J. W. (2000). A meta-analysis of the effectiveness of telecourses in distance education. *American Journal of Distance Education, 14*(1), 27–46.

*Magiera, F. T. (1994). Teaching managerial finance through compressed video: An alternative for distance education. *Journal of Education for Business, 69*(5), 273–277.

*McFall, J. P., & Freddolino, P. P. (2000). Quality and comparability in distance field education: Lessons learned from comparing three program sites. *Journal of Social Work Education, 36*(2), 293–307.

*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.

McIsaac, M. S., & Gunawardena, C. N. (1996). Distance education. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology: A project of the Association for Educational Communications and Technology* (pp. 403–437). New York: Simon & Schuster Macmillan.

*Miller, G., & Pilcher, C. L. (2001). Levels of cognition reached in agricultural distance education courses in comparison to on-campus courses and to faculty perceptions concerning an appropriate level. *Journal of Agricultural Education, 42*(1), 20–27.

*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.

Moore, M. G., & Thompson, M. M. (1996). *The effects of distance learning* (Rev. ed.; Research Monograph No. 15). University Park: Pennsylvania State University, American Center for the Study of Distance Education.

*Murphy, T. H. (2000). An evaluation of a distance education course design for general soils. *Journal of Agricultural Education, 41*(3), 103–113.

National Center for Education Statistics. (1999). Distance education at postsecondary education institutions, 1997–98 (Statistical analysis report, NCES No. 2000-013). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved May 23, 2005, from http://nces.ed.gov/pubs2000/2000013.pdf

National Center for Education Statistics. (2003). *Distance education at degree-granting postsecondary institutions: 2000–2001* (NCES No. 2003-017). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

*Nesler, M. S., Hanner, M. B., & Melburg, V. (2001). Professional socialization of baccalaureate nursing students: Can students in distance nursing programs become socialized? *Journal of Nursing Education, 40*(7), 293–302.

*Phelps, R. H., Wells, R. A., Ashworth, R. L., Jr., & Heidi, A. H. (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*(6), 351–356.

*Redding, T. R., & Rotzien, J. (2001). Comparative analysis of online learning versus classroom learning. *Journal of Interactive Instruction Development, 13*(4), 3–12.

*Roblyer, M. D. (1999). Is choice important in distance learning? A study of student motives for taking Internet-based courses at the high school and community college levels. *Journal of Research on Computing in Education, 32*(1), 157–171.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* Newbury Park, CA: Sage.

Russell, T. L. (1999). *The no significant difference phenomenon.* Chapel Hill: North Carolina State University, Office of Instructional Telecommunications.

*Ryan, M., Carlton, K. H., & Ali, N. S. (1999). Evaluation of traditional classroom teaching methods versus course delivery via the World Wide Web. *Journal of Nursing Education, 38*(6), 272–277.

*Ryan, W. F. (1996). The effectiveness of traditional vs. audiographics delivery in senior high advanced mathematics courses. *Journal of Distance Education, 11*(2), 45–55.

Schlosser, C. A., & Anderson, M. L. (1994). *Distance education: Review of the literature.* Ames, IA: Iowa Distance Education Alliance. (ERIC Document Reproduction Service No. ED382159)

*Schoenfeld-Tacher, R., McConnell, S., & Gra, M. (2001). Do no harm—A comparison of the effects of on-line vs. traditional delivery media on a science course. *Journal of Science Education and Technology, 10*(3), 257–265.

Schramm, W. (1962). Learning from instructional television. *Review of Educational Research, 32*(2), 156–167.

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

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In *The handbook of research synthesis* (pp. 261–284). New York: Russell Sage Foundation.

Shale, D. (1990). Toward a reconceptualization of distance education. In M. G. Moore (Ed.), *Contemporary issues in American distance education* (pp. 333–343). Oxford, UK: Pergamon Press.

*Simpson, H., Pugh, H. L., & Parchman, S. W. (1991). An experimental two-way video teletraining system: Design, development and evaluation. *Distance Education, 12*(2), 209–231.

*Smeaton, A. F., & Keogh, G. (1999). An analysis of the use of virtual delivery of undergraduate lectures. *Computers & Education, 32*(1), 83–94.

*Smith, S. B., Smith, S. J., & Boone, R. (2000). Increasing access to teacher preparation: The effectiveness of traditional instructional methods in an online learning environment. *Journal of Special Education Technology, 15*(2), 37–46.

*Sonner, B. S. (1999). Success in the capstone business course—Assessing the effectiveness of distance learning. *Journal of Education for Business, 74*(4), 243–247.

*Sowinski, K. M., Scott, S. A., & Carlstedt, B. C. (2000). An Internet training module for traditional and nontraditional doctor of pharmacy students: Content and evaluation. *American Journal of Pharmaceutical Education, 64*(1), 15–19.

*Spooner, F., Jordan, L., Algozzine, B., & Spo, M. (1999). Student ratings of instruction in distance learning and on-campus classes. *Journal of Educational Research, 92*(3), 132–140.
*Steyn, M. M. d. V., Alexander, P. M., & Röhm, D. (1996). CAL for first year analytical chemistry by distance education. *Computers & Education, 27*(2), 95–101.
Stickell, D. W. (1963). *A critical review of the methodology and results of research comparing televised and face-to-face instruction.* University Park: Pennsylvania State University.
Stock, W. A. (1994). Systematic coding for research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 125–138). New York: Russell Sage Foundation.
*Sugrue, B., Rietz, T., & Hansen, S. (1999). Distance learning: Relationships among class size, instructor location, student perceptions, and performance. *Performance Improvement Quarterly, 12*(3), 44–57.
*Summers, M., Anderson, J. L., Lines, A. R., Gelder, B. C., & Dean, R. S. (1996). The camera adds more than pounds: Gender differences in course satisfaction for campus and distance learning students. *Journal of Research and Development in Education, 29*(4), 212–219.
*Thyer, B. A., Artelt, T., & Markward, M. (1998). Evaluating distance learning in social work education: A replication study. *Journal of Social Work Education, 34*(2), 291–295.
Twigg, C. A. (2001). *Innovations in online learning: Moving beyond no significant difference.* Taylor, NY: Pew Learning and Technology.
United States General Accounting Office. (2002). *Growth in distance education programs and implications for federal education policy: Statement of Cornelia M. Ashby, Director.* Washington, DC: Author.
*Wagner, N., & Craft, E. H. (1988). Instructional television fixed service in Arizona. *American Journal of Distance Education, 2*(1), 76–80.
*Wang, A. Y., & Newlin, M. H. (2000). Characteristics of students who enroll and succeed in psychology Web-based classes. *Journal of Educational Psychology, 92*(1), 137–143.
*Wilkinson, K. L., & Hemby, K. V. (2000). An examination of perceptions of the use of virtual conferences in organizations: The Organizational Systems Research Association (OSRA) and the Association for Business Communication (ABC) Members. *Information Technology, Learning, and Performance Journal, 18*(2), 13–23.
*Wisher, R. A., & Curnow, C. K. (1999). Perceptions and effects of image transmissions during Internet-based training. *American Journal of Distance Education, 13*(3), 37–51.

YONG ZHAO is University Distinguished Professor in the Learning, Technology, and Culture Program in the College of Education at Michigan State University. His research interests include educational uses of technology, online/distance education, and innovation diffusions. His recent publications include "Technology Uses in Schools: An Ecological Perspective" (*American Educational Research Journal,* 2003, with K. Frank) and an edited volume, *What Should Teachers Know about Technology: Perspectives and Practices* (Information Age, 2003).

JING LEI is an assistant professor in the Instructional Design Development and Evaluation program in the School of Education at Syracuse University. Her research interests include educational technology integration, meaningful technology use in schools, socio-cultural and psychological impact of technology, teacher educational technology professional development, and international and comparative education.

BO YAN is a doctoral student in the Learning, Technology, and Culture Program in the College of Education at Michigan State University.

CHUN LAI is a doctoral student in the Learning, Technology, and Culture Program in the College of Education at Michigan State University.

SOPHIA TAN is an assistant professor in the Professional Program in Teacher Education at Coastal South Carolina University.