Meta-analysis is a method of quantitatively summarizing the results of experimental research. This article summarizes four meta-analyses published since 2003 that compare the effect of DE and traditional education (TE) on student learning. Despite limitations, synthesis of these meta-analyses establish, at the very least, equivalent learning outcomes for DE and TE. Research efforts should now be directed toward the more important question: Why is DE not clearly superior to TE? TE is restricted by time-place-pace: DE is marginally influenced by such restrictions. Intuitively, learning opportunities that are convenient and individualized, as opposed to those that are fixed and inflexible, should result in higher student achievement. DE demands learner responsibility and affords learner autonomy, both of which facilitate academic achievement. Instructional technology, virtually synonymous with DE, is a force that can transform instruction. Why is DE not clearly superior to TE?

Distance education (DE) is “planned learning that normally occurs in a different place from teaching and as a result requires special techniques of course design, special instructional techniques, special methods of communication by electronic and other technology, as well as special organizational and administrative arrangements” (Moore & Kearsley, 1996, p. 2). Holmborg (1989) claimed that DE is “characterized by non-contiguous communication and can be carried out anywhere and at any time” (p. 168). Rumble
(1989), however, suggested that “educational and training systems all fall somewhere on a continuum that ranges from the purely contiguous to the purely distant” (p. 35). In this regard, it may be useful to conceptualize instructional format in terms of sensitivity to time, place, and pace (Fletcher, 2001; Shachar & Neumann, 2003). How does instruction restricted by time, place, and pace compare to instruction devoid of such restrictions? “The answer to the DE effectiveness question, or any research question for that matter, cannot be found in a single study” (Bernard et al., 2004, p. 383).

META-ANALYSIS AND DISTANCE EDUCATION

Meta-analysis is a type of quantitative synthesis that allows for collective interpretation of many research studies by extracting an effect size from each study (Glass, McGaw, & Smith, 1981). Effect size is the average difference between treatment and control groups in standard form (Shachar & Neumann, 2003). As opposed to reviews of the literature, meta-analysis provides a “sense of the possible impact of a procedure against all published studies, rather than an illustrative few selected by the reviewer” (Heberlein & Baumgartner, 1978, p. 448). Effect size values of 0.2, 0.5, and 0.8 are considered small, medium, and large, respectively (Cohen, 1988).

This article summarizes four recent meta-analyses that compare the effectiveness of DE with traditional education (TE). While educational access and student satisfaction are important criteria by which to evaluate course delivery formats (Allen et al., 2002), learning is the ultimate measure of instructional effectiveness (Bernard et al., 2004). Therefore, only meta-analyses that quantify student learning outcomes are reviewed.

DIFFERENCES BETWEEN TRADITIONAL AND DISTANCE EDUCATION

ACADEMIC PERFORMANCE

In 2003, Shachar and Neumann ((2003) conducted a meta-analysis to answer the research question: “Is there a difference in the Final Academic Performance of students enrolled in DE programs compared to those enrolled in traditional F2F programs?” (p. 10). DE was not defined and studies were included in the analysis on the basis of unspecified search
criteria. Eighty-six experimental and quasi-experimental studies, published between 1990 and 2002, written in English or with an English translation, with no severe methodological flaws, were included in the meta-analysis. Data extraction from these studies results in 86 effect sizes for final academic performance. Aggregating all studies resulted in a collective sample of over 15,300 students: 7,270 in DE and 8,076 in TE.

In approximately “two thirds of the cases, students taking courses by distance education outperformed their student counterparts enrolled in traditionally instructed courses” (Shachar & Neumann, 2003, p. 1). Computation of effect size yielded the final result of 0.37, with a 95% confidence interval of 0.33 to 0.40 (i.e., 95 times out of 100, effect size falls within the confidence interval). Such an effect size indicates that a student achieving at the 50th percentile in TE would achieve at the 65th percentile had they received DE. The researchers addressed the file draw problem (i.e., studies that do not report significant differences between groups are more likely to be filed than published). Approximately 3,062 additional unreported studies that found no difference between treatment and control would be necessary to nullify this effect size. The researchers suggested that “it was unlikely that there were that many well constructed studies sitting in file drawers” (p. 11).

Limitations of the Shachar and Neumann (2003) meta-analysis warrant cautious interpretation of their findings. First, there is no report of the proportion of quasi-experimental studies included in the meta-analysis (i.e., used preformed groups). True experiments in DE effectiveness research are rare. It is difficult to imagine a situation in which a large group of students volunteer for random assignment to DE or TE. It is reasonable to infer that most of the 86 comparisons were between students who chose DE and those who chose TE. Sherry (1995) reported that DE students were more mature and more motivated than TE students. Thus, we cannot be certain if DE students would have out-performed the control group had the treatment not occurred. Second, the researchers are vague about the outcome measure. Final academic performance may refer to teacher assigned grades. Such a measure of student learning is influenced by teacher characteristics (Bone-ronning, 1999) and DE teachers are different than TE teachers (Mielke, 1999). It may be that DE teachers assign higher marks than TE teachers. In this regard, it may be differences in teachers, not differences in mode of delivery, that explain differences in achievement between DE and TE students. Finally, given that a precise definition of DE did not guide study selection, it is not clear which educational formats were compared.
HOW DOES DISTANCE EDUCATION COMPARE WITH CLASSROOM INSTRUCTION?

In 2004, Bernard and his colleagues conducted a meta-analysis of 232 studies published since 1985 that compared the effects of DE and TE on student achievement, attitude, and retention. “The year 1985 was chosen as a cutoff date because electronically mediated, interactive DE became widely available around this time” (p. 388). DE was defined as: (a) the semi-permanent separation (place and/or time) of learner and instructor; (b) presence of planning, preparation, and student support; (c) provision of two-way media to facilitate dialog and interaction between student and teacher and among students. Achievement outcomes included objective measures (i.e., standardized tests, researcher-made tests, teacher-made tests) that assessed the extent to which students had mastered the learning objectives of the course. To facilitate meaningful comparison, three research study features were coded: methodology (e.g., equality between treatment and control groups, extent of missing data, selection bias), pedagogy (e.g., systematic instructional design, advanced course information given to students, contact with teacher), and media (e.g., computer-based instruction, teleconferencing, Internet). The 232 studies yielded 321 achievement effect sizes for a total of 57,019 students.

Results suggested that methodology was a good predictor of student achievement; methods used to study achievement under different instructional conditions (i.e., DE and TE) were predictive of achievement outcomes. Pedagogy predicted student achievement; in both DE and TE, sound instructional practice facilitated student learning. Instructional practice was more important to student achievement than the inclusion of media in the learning process. In terms of overall achievement, there was a small but significant effect favoring DE ($d = 0.013$). Further analysis revealed that in mathematics, science, and engineering courses, TE students out-performed DE students. In courses related to computing, military, and business, DE students out-performed TE students. Bernard and his colleagues (2004) noted wide variability in effect sizes for all comparisons. Such variability suggested that a substantial number of DE applications resulted in better student achievement than TE. “On the other hand, a substantial number of DE applications are far worse than classroom instruction” (p. 346).

The Bernard et al. (2004) meta-analysis is impressive in its rigor, although an element of caution is necessary in interpreting results. While the researchers stated that only studies that used objective measures of student
achievement were included in the analysis, teacher-made tests and, particularly, teacher marking of those tests are not necessarily objective (Bonesronning, 1999). DE and TE teachers may not be consistent in examination use and/or marking; differences in student grades may be due to teacher differences rather than student learning differences. Additionally, Bernard did not address the file draw problem. As was the case with the Shachar and Neumann (2003) meta-analysis, Bernard et al. provided no indication of the proportion of studies that were quasi-experiential (i.e., used preformed groups). It may be that few, if any, of the 232 studies were true experiments. However, given that many variables were coded and controlled, the impact of extraneous influences on student learning is reduced but not likely eliminated.

**EVALUATING THE EFFECTIVENESS OF DISTANCE LEARNING**

In 2004, Allen and his colleagues conducted a meta-analysis of 38 studies that compared DE and TE on “at least one assessment of student performance in the course related to mastery of some content or skill taught in the course” (p. 408). DE was not defined and research studies were identified by the search terms distance learning or distance education. Date of publication was not a criterion for study selection and research published in the 1970s was included in the analysis. Three potentially moderating variables were coding in order to examine possible influences on student achievement in DE. DE was coded as synchronous or asynchronous. To be designated as synchronous, communication between teacher and student was simultaneous (e.g., two-way audiovisual links between two or more environments). Conversely, an asynchronous system was one in which students could not communicate with the teacher in real time (e.g., e-mail). DE was also coded in terms of channel of delivery (i.e., video, audio, written text) and course content (i.e., sciences, military training, foreign languages, social sciences, education).

“The average effect demonstrated a small improvement in performance for courses using distance education technology (average $r = .048$)” (Allen et al., 2004, p. 410). The sample of correlations was heterogeneous suggesting cautious interpretation of average effect size. A longitudinal study of Tanzanian teacher-training programs was deleted from analyses of moderator variables due to the influence of large sample size (i.e., 63,516 students).
Synchronous or asynchronous DE did not differentially impact student achievement. The use of video and written text as channels of instruction was associated with slightly better DE outcomes than audio methods of course delivery. With regard to course content, DE lowered student performance in military-related instruction, increased achievement in foreign language learning, and had no effect on achievement in the remaining content areas.

The Allen et al. (2004) meta-analysis has limitations that compromise interpretation of results. Given the evolution of DE technologies, combining research conducted in the 1970s with research conducted after 2000 does not result in meaningful analysis. As was the case with Shachar and Neumann (2003) and Bernard et al. (2004), while treatment and control groups were required for inclusion in meta-analyses, true experiments were, at the very least, underrepresented relative to quasi-experiments. Indeed, in the Allen et al. meta-analysis, the objectivity of the outcome measures was questionable. The equivalence of teacher competencies across instructional conditions (i.e., DE and TE) was not established and the file draw problem was not addressed. Finally, given that a precise definition of DE did not guide study selection, it is not clear which educational formats were compared.
amount of teacher training in DE); and methodology (e.g., type of achievement measure).

“The analysis resulted in an overall weighted effect size not significantly different from zero, a result that is consistent with the results of recent meta-analyses of distance education” (Cavanaugh et al., 2004, p. 16). Effect sizes ranged from -1.158 to 0.597, “indicating that some applications of distance education appeared to be much better than classroom instruction and others were much worse” (p. 16). All but one of the 116 outcomes had effect sizes that were not significant, suggesting that DE students consistently performed as well as students in classroom-based programs. No coded study features, including instructional and program variables, increased DE effect sizes.

As was the case with Shachar and Neumann (2003), Bernard et al. (2004), and Allen et al. (2004), caution is warranted in interpreting the results of the Cavanaugh et al. (2004) meta-analysis. Random assignment of students to DE or TE remains a concern. Children who participate in web-based telecommunication educational programs are likely a unique group. However, coding for research design characteristics decreased the impact of non-random assignment. As well, curriculum and assessment standardization in the context of public education suggests increased probability of equivalency across DE and TE conditions. Given the consistency of the effects reported in recent empirical literature, Cavanaugh et al. concluded that, “educators and other stakeholders can reasonably expect learning in a well-designed distance education environment to be equivalent to learning in a well-designed classroom environment” (p. 20).

Results of the four recent meta-analyses on the relative effectiveness of DE and TE are summarized and presented in Table 1. While caution in interpreting results is required, a conclusion emerges.
Table 1
Summary of Meta-Analysis

<table>
<thead>
<tr>
<th>Meta-Analysis</th>
<th>Studies</th>
<th>Result</th>
<th>Effectiveness of Distance Education (DE) versus Traditional Education (TE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shachar &amp; Neumann</td>
<td>86</td>
<td>$d = 0.37$</td>
<td>While there is wide variability in research findings on the relative</td>
</tr>
<tr>
<td>(2003)</td>
<td></td>
<td>0.00 to 0.40</td>
<td>effectiveness of DE and TE, in general, DE results in mildly superior</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>student learning outcomes.</td>
</tr>
<tr>
<td>Bernard et al.</td>
<td>232</td>
<td>$d = 0.01$</td>
<td>While there is wide variability in research findings on the relative</td>
</tr>
<tr>
<td>(2004)</td>
<td></td>
<td>-1.31 to 1.41</td>
<td>effectiveness of DE and TE, in general, DE results in slightly superior</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>student learning outcomes.</td>
</tr>
<tr>
<td>Allen et al.</td>
<td>38</td>
<td>$r = .048$</td>
<td>While there is wide variability in research findings on the relative</td>
</tr>
<tr>
<td>(2004)</td>
<td></td>
<td>heterogeneous</td>
<td>effectiveness of DE and TE, in general, DE results in mildly superior</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>student learning outcomes.</td>
</tr>
<tr>
<td>Cavanaugh et al.</td>
<td>14</td>
<td>$d = -0.03$</td>
<td>While there is wide variability in research findings on the relative</td>
</tr>
<tr>
<td>(2004)</td>
<td></td>
<td>-0.47 to 0.01</td>
<td>effectiveness of DE and TE, in general, there is no difference in</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>student learning outcomes.</td>
</tr>
</tbody>
</table>
CONCLUSION: THE EVOLUTION OF INSTRUCTION AND THE LIMITS OF RESEARCH

People tend to conceptualize emerging technology in relation to existing technology; “the automobile was called the horseless carriage and the railroad train was called the iron horse” (Bernard et al., 2004, p. 379). Often accompanied by apprehension, such comparisons result in prudent implementation of innovation. Just as the term horseless carriage was temporary and transitional, so too may be the term distance education. Relative to the early automobile, the horse appeared superior. Lack of suitable roads, engine vulnerability, inadequate safety precautions, and no driver training resulted in chronic problems; the horse and rider frequently plodded by the stranded automobile and driver.

While “contradictory findings continue to point to the proposition that a complete meta-analysis is required” (Allen et al., 2004, p. 406), recent meta-analyses have not entirely resolved the effectiveness debate. It is not the meta-analytic method that falls short but, rather, the primary research upon which meta-analyses are based. There is consensus that, “in general, DE research is of a low quality, particularly in terms of internal validity (i.e., control for confounds and inequalities)” (Bernard et al., 2004, p. 416). In evaluating the effectiveness of DE, it is difficult to conduct true experiments and easy to conduct quasi-experiments. Because initial (i.e., prior to DE instruction) group equivalence is not established or ensured (i.e., by random assignment of students to DE or TE), subsequent differences in student achievement (i.e., after DE instruction) cannot be interpreted with certainty.

The ideal research design for determination of the relative effectiveness of DE and TE is costly and complex. Random assignment of teachers to DE or TE courses would ensure equivalent instructional competencies across conditions. Correspondingly, random assignment of students to DE or TE courses would ensure equivalent student characteristics across conditions. Alternatively, teachers and students in DE and TE could be systematically matched to ensure that instructional format was the only difference between groups. Blomeyer (2002) argued that “scientific educational research should be given a high priority in all public and private agencies supporting implementation and use of online learning” (p. 10). And yet, “the demand for research will always lag behind the supply of research, and for this reason it is important to apportion our collective research resources judiciously” (Bernard et al., 2004, p. 416).
It could be argued that extreme application of scientific rigor is unnecessary in evaluating the relative effectiveness of DE and TE. Efforts are often made in meta-analyses of DE effectiveness research to code study features (e.g., instructional variables) in order to control for extraneous factors (Allen et al., 2004; Bernard et al., 2004; Cavanaugh et al., 2004) and to apply statistical analyses that contribute to increased certainty in interpretation of findings (Shachar & Neumann, 2003). Nonetheless, despite limitations, the magnitude of primary research in the area as well as numerous meta-analyses strongly suggest, at the very least, equivalent learning outcomes for DE and TE. If this conclusion is accepted, research effort can be directed toward answering the more important question: Why is DE not clearly superior to TE?

TE is controlled by time-place-pace: DE is marginally influenced by such control. Intuitively, learning opportunities that are convenient and individualized, as opposed to those that are fixed and inflexible, should result in higher student achievement. Relative to TE, DE demands learner responsibility and affords learner autonomy (Keegan, 1996), both of which facilitate academic achievement (Derrick & Carr, 2003; Hiemstra, 1994). Instructional technology, virtually synonymous with DE, is “a force that can transform education because of the power of e-learning to individualize, personalize and differentiate instruction” (Cavanaugh et al., 2004, p. 23). Why is DE not clearly superior to TE?

The automobile and corresponding transportation infrastructure evolved; the horse and carriage disappeared to all but the nostalgic. TE is in a process of transformation as “advancements in increasingly flexible technology have enabled the Web’s visual, interactive nature to transform the traditional campus classroom-instructor system into a variety of different and innovative forms of instructional dissemination and to decentralized locations” (Shachar & Neumann, 2003, p. 3). In the near future, if not currently, the distinction between DE and TE may be an artifact that has outgrown educational utility. The legacy of DE and TE may be the dynamic tension created by their juxtaposition, a tension that advanced the only legitimate educational agenda—maximizing the learning of every student and maximizing the number of students served.
References


approach. *International Review of Research in Open and Distance Education*. Retrieved February 6, 2005, from http://www.irrodl.org/content/v4.2/shachar-neumann.html