U.S. DEPARTMENT OF EDUCATION

Evaluation of Evidence-Based Practices in Online Learning

A Meta-Analysis and Review of Online Learning Studies

Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies

U.S. Department of Education
Office of Planning, Evaluation, and Policy Development
Policy and Program Studies Service

Revised September 2010

Prepared by

Barbara Means
Yukie Toyama
Robert Murphy
Marianne Bakia
Karla Jones

Center for Technology in Learning

This report was prepared for the U.S. Department of Education under Contract number ED-04-CO-0040 Task 0006 with SRI International. Bernadette Adams Yates served as the project manager. The views expressed herein do not necessarily represent the positions or policies of the Department of Education. No official endorsement by the U.S. Department of Education is intended or should be inferred.

U.S. Department of Education
Arne Duncan
Secretary

Office of Planning, Evaluation and Policy Development
Carmel Martin
Assistant Secretary

Policy and Program Studies Service
Alan Ginsburg
Director

Office of Educational Technology
Karen Cator
Director

September 2010

This report is in the public domain. Authorization to reproduce this report in whole or in part is granted. Although permission to reprint this publication is not necessary, the suggested citation is: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development, *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies,* Washington, D.C., 2010.

This report is also available on the Department's Web site at www.ed.gov/about/offices/list/opepd/ppss/reports.html.

On request, this publication is available in alternate formats, such as braille, large print, or computer diskette. For more information, please contact the Department's Alternate Format Center at (202) 260-0852 or (202) 260-0818.

Contents

EXHIBITS v

ACKNOWLEDGMENTS vii

ABSTRACT ix

EXECUTIVE SUMMARY xi

Literature Search xii
Meta-Analysis xiii
Narrative Synthesis xiv
Key Findings xiv
Conclusions xviii

1. INTRODUCTION 1

Context for the Meta-analysis and Literature Review 2
Conceptual Framework for Online Learning 3
Findings From Prior Meta-Analyses 6
Structure of the Report 7

2. METHODOLOGY 9

Definition of Online Learning 9
Data Sources and Search Strategies 10
Electronic Database Searches 10
Additional Search Activities 10
Screening Process 11
Effect Size Extraction 13
Coding of Study Features 14
Data Analysis 15

3. FINDINGS 17

Nature of the Studies in the Meta-Analysis 17
Main Effects 18
Test for Homogeneity 27
Analyses of Moderator Variables 27
Practice Variables 28
Condition Variables 30
Methods Variables 31

4. NARRATIVE SYNTHESIS OF STUDIES COMPARING VARIANTS OF ONLINE LEARNING 37

Blended Compared With Pure Online Learning 38
Media Elements 40
Learning Experience Type 41
Computer-Based Instruction 43
Supports for Learner Reflection 44
Moderating Online Groups 46
Scripts for Online Interaction 46
Delivery Platform 47
Summary 48

5. DISCUSSION AND IMPLICATIONS 51

Comparison With Meta-Analyses of Distance Learning 52
Implications for K–12 Education 53

REFERENCES **55**

Reference Key 55

APPENDIX META-ANALYSIS METHODOLOGY **A-1**

Terms and Processes Used in the Database Searches A-1
Additional Sources of Articles A-3
Effect Size Extraction A-4
Coding of Study Features A-5

Exhibits

Exhibit 1. Conceptual Framework for Online Learning 5
Exhibit 2. Bases for Excluding Studies During the Full-Text Screening Process 13
Exhibit 3. Effect Sizes for Contrasts in the Meta-Analysis 20
Exhibit 4a. Purely Online Versus Face-to-Face (Category 1) Studies Included in the Meta-Analysis 21
Exhibit 4b. Blended Versus Face-to-Face (Category 2) Studies Included in the Meta-Analysis 24
Exhibit 5. Tests of Practices as Moderator Variables 29
Exhibit 6. Tests of Conditions as Moderator Variables 30
Exhibit 7. Studies of Online Learning Involving K–12 Students 32
Exhibit 8. Tests of Study Features as Moderator Variables 34
Exhibit 9. Learner Types for Category 3 Studies 37
Exhibit A-1. Terms for Initial Research Database Search A-2
Exhibit A-2. Terms for Additional Database Searches for Online Career Technical Education and Teacher Professional Development A-2
Exhibit A-3. Sources for Articles in the Full-Text Screening A-3
Exhibit A-4. Top-level Coding Structure for the Meta-analysis A-6

Acknowledgments

This revision to the 2009 version of the report contains corrections made after the discovery of several transcription errors by Shanna Smith Jaggars and Thomas Bailey of the Community College Research Center of Teachers College, Columbia University. We are indebted to Jaggars and Bailey for their detailed review of the analysis.

We would like to acknowledge the thoughtful contributions of the members of our Technical Work Group in reviewing study materials and prioritizing issues to investigate. The advisors consisted of Robert M. Bernard of Concordia University, Richard E. Clark of the University of Southern California, Barry Fishman of the University of Michigan, Dexter Fletcher of the Institute for Defense Analysis, Karen Johnson of the Minnesota Department of Education, Mary Kadera of PBS, James L. Morrison an independent consultant, Susan Patrick of the North American Council for Online Learning, Kurt D. Squire of the University of Wisconsin, Bill Thomas of the Southern Regional Education Board, Bob Tinker of The Concord Consortium, and Julie Young of the Florida Virtual School. Robert M. Bernard, the Technical Work Group's meta-analysis expert, deserves a special thanks for his advice and sharing of unpublished work on meta-analysis methodology as well as his careful review of an earlier version of this report.

Many U.S. Department of Education staff members contributed to the completion of this report. Bernadette Adams Yates served as project manager and provided valuable substantive guidance and support throughout the design, implementation and reporting phases of this study. We would also like to acknowledge the assistance of other Department staff members in reviewing this report and providing useful comments and suggestions, including David Goodwin, Daphne Kaplan, Tim Magner, and Ze'ev Wurman.

We appreciate the assistance and support of all of the above individuals; any errors in judgment or fact are of course the responsibility of the authors.

The Study of Education Data Systems and Decision Making was supported by a large project team at SRI International. Among the staff members who contributed to the research were Sarah Bardack, Ruchi Bhanot, Kate Borelli, Sara Carriere, Katherine Ferguson, Reina Fujii, Joanne Hawkins, Ann House, Katie Kaattari, Klaus Krause, Yessica Lopez, Lucy Ludwig, Patrik Lundh, L. Nguyen, Julie Remold, Elizabeth Rivera, Luisana Sahagun Velasco, Mark Schlager, and Edith Yang.

Abstract

A systematic search of the research literature from 1996 through July 2008 identified more than a thousand empirical studies of online learning. Analysts screened these studies to find those that (a) contrasted an online to a face-to-face condition, (b) measured student learning outcomes, (c) used a rigorous research design, and (d) provided adequate information to calculate an effect size. As a result of this screening, 50 independent effects were identified that could be subjected to meta-analysis. The meta-analysis found that, on average, students in online learning conditions performed modestly better than those receiving face-to-face instruction. The difference between student outcomes for online and face-to-face classes—measured as the difference between treatment and control means, divided by the pooled standard deviation—was larger in those studies contrasting conditions that blended elements of online and face-to-face instruction with conditions taught entirely face-to-face. Analysts noted that these blended conditions often included additional learning time and instructional elements not received by students in control conditions. This finding suggests that the positive effects associated with blended learning should not be attributed to the media, per se. An unexpected finding was the small number of rigorous published studies contrasting online and face-to-face learning conditions for K–12 students. In light of this small corpus, caution is required in generalizing to the K–12 population because the results are derived for the most part from studies in other settings (e.g., medical training, higher education).

Executive Summary

Online learning—for students and for teachers—is one of the fastest growing trends in educational uses of technology. The National Center for Education Statistics (2008) estimated that the number of K-12 public school students enrolling in a technology-based distance education course grew by 65 percent in the two years from 2002-03 to 2004-05. On the basis of a more recent district survey, Picciano and Seaman (2009) estimated that more than a million K–12 students took online courses in school year 2007–08.

Online learning overlaps with the broader category of distance learning, which encompasses earlier technologies such as correspondence courses, educational television and videoconferencing. Earlier studies of distance learning concluded that these technologies were not significantly different from regular classroom learning in terms of effectiveness. Policy-makers reasoned that if online instruction is no worse than traditional instruction in terms of student outcomes, then online education initiatives could be justified on the basis of cost efficiency or need to provide access to learners in settings where face-to-face instruction is not feasible. The question of the relative efficacy of online and face-to-face instruction needs to be revisited, however, in light of today's online learning applications, which can take advantage of a wide range of Web resources, including not only multimedia but also Web-based applications and new collaboration technologies. These forms of online learning are a far cry from the televised broadcasts and videoconferencing that characterized earlier generations of distance education. Moreover, interest in hybrid approaches that blend in-class and online activities is increasing. Policy-makers and practitioners want to know about the effectiveness of Internet-based, interactive online learning approaches and need information about the conditions under which online learning is effective.

The findings presented here are derived from (a) a systematic search for empirical studies of the effectiveness of online learning and (b) a meta-analysis of those studies from which effect sizes that contrasted online and face-to-face instruction could be extracted or estimated. A narrative summary of studies comparing different forms of online learning is also provided.

These activities were undertaken to address four research questions:

1. *How does the effectiveness of online learning compare with that of face-to-face instruction?*
2. *Does supplementing face-to-face instruction with online instruction enhance learning?*
3. *What practices are associated with more effective online learning?*
4. *What conditions influence the effectiveness of online learning?*

This meta-analysis and review of empirical online learning research are part of a broader study of practices in online learning being conducted by SRI International for the Policy and Program Studies Service of the U.S. Department of Education. The goal of the study as a whole is to provide policy-makers, administrators and educators with research-based guidance about how to implement online learning for K–12 education and teacher preparation. An unexpected finding of the literature search, however, was the small number of published studies contrasting online and

face-to-face learning conditions for K–12 students. Because the search encompassed the research literature not only on K–12 education but also on career technology, medical and higher education, as well as corporate and military training, it yielded enough studies with older learners to justify a quantitative meta-analysis. Thus, analytic findings with implications for K–12 learning are reported here, but caution is required in generalizing to the K–12 population because the results are derived for the most part from studies in other settings (e.g., medical training, higher education).

This literature review and meta-analysis differ from recent meta-analyses of distance learning in that they

- Limit the search to studies of Web-based instruction (i.e., eliminating studies of video- and audio-based telecourses or stand-alone, computer-based instruction);
- Include only studies with random-assignment or controlled quasi-experimental designs; and
- Examine effects only for objective measures of student learning (e.g., discarding effects for student or teacher perceptions of learning or course quality, student affect, etc.).

This analysis and review distinguish between instruction that is offered entirely online and instruction that combines online and face-to-face elements. The first of the alternatives to classroom-based instruction, entirely online instruction, is attractive on the basis of cost and convenience as long as it is as effective as classroom instruction. The second alternative, which the online learning field generally refers to as blended or hybrid learning, needs to be *more* effective than conventional face-to-face instruction to justify the additional time and costs it entails. Because the evaluation criteria for the two types of learning differ, this meta-analysis presents separate estimates of mean effect size for the two subsets of studies.

Literature Search

The most unexpected finding was that an extensive initial search of the published literature from 1996 through 2006 found *no* experimental or controlled quasi-experimental studies that both compared the learning effectiveness of online and face-to-face instruction for K–12 students and provided sufficient data for inclusion in a meta-analysis. A subsequent search extended the time frame for studies through July 2008.

The computerized searches of online databases and citations in prior meta-analyses of distance learning as well as a manual search of the last three years of key journals returned 1,132 abstracts. In two stages of screening of the abstracts and full texts of the articles, 176 online learning research studies published between 1996 and 2008 were identified that used an experimental or quasi-experimental design and objectively measured student learning outcomes. Of these 176 studies, 99 had at least one contrast between an included online or blended learning condition and face-to-face (offline) instruction that potentially could be used in the quantitative meta-analysis. Just nine of these 99 involved K–12 learners. The 77 studies without a face-to-face condition compared different variations of online learning (without a face-to-face control condition) and were set aside for narrative synthesis.

Meta-Analysis

Meta-analysis is a technique for combining the results of multiple experiments or quasi-experiments to obtain a composite estimate of the size of the effect. The result of each experiment is expressed as an *effect size*, which is the difference between the mean for the treatment group and the mean for the control group, divided by the pooled standard deviation. Of the 99 studies comparing online and face-to-face conditions, 45 provided sufficient data to compute or estimate 50 independent effect sizes (some studies included more than one effect). Four of the nine studies involving K–12 learners were excluded from the meta-analysis: Two were quasi-experiments without statistical control for preexisting group differences; the other two failed to provide sufficient information to support computation of an effect size.

Most of the articles containing the 50 effects in the meta-analysis were published in 2004 or more recently. The split between studies of purely online learning and those contrasting blended online/face-to-face conditions against face-to-face instruction was fairly even, with 27 effects in the first category and 23 in the second. The 50 estimated effect sizes included seven contrasts from five studies conducted with K–12 learners—two from eighth-grade students in social studies classes, one for eighth- and ninth-grade students taking Algebra I, two from a study of middle school students taking Spanish, one for fifth-grade students in science classes in Taiwan, and one from elementary-age students in special education classes. The types of learners in the remaining studies were about evenly split between college or community college students and graduate students or adults receiving professional training. All but two of the studies involved formal instruction. The most common subject matter was medicine or health care. Other content types were computer science, teacher education, mathematics, languages, science, social science, and business. Among the 48 contrasts from studies that indicated the time period over which instruction occurred, 19 involved instructional time frames of less than a month, and the remainder involved longer periods. In terms of instructional features, the online learning conditions in these studies were less likely to be instructor-directed (8 contrasts) than they were to be student-directed, independent learning (17 contrasts) or interactive and collaborative in nature (22 contrasts).

Effect sizes were computed or estimated for this final set of 50 contrasts. Among the 50 individual study effects, 11 were significantly positive, favoring the online or blended learning condition. Three contrasts found a statistically significant effect favoring the traditional face-to-face condition.[1]

[1] When a $\alpha < .05$ level of significance is used for contrasts, one would expect approximately 1 in 20 contrasts to show a significant difference by chance. For 50 contrasts, then, one would expect 2 or 3 significant differences by chance. The finding of 3 significant contrasts associated with face-to-face instruction is within the range one would expect by chance; the 11 contrasts associated with online or hybrid instruction exceeds what one would expect by chance.

Narrative Synthesis

In addition to the meta-analysis comparing online learning conditions with face-to-face instruction, analysts reviewed and summarized experimental and quasi-experimental studies contrasting different versions of online learning. Some of these studies contrasted purely online learning conditions with classes that combined online and face-to-face interactions. Others explored online learning with and without elements such as video, online quizzes, assigned groups, or guidance for online activities. Five of these studies involved K–12 learners.

Key Findings

The main finding from the literature review was that

- *Few rigorous research studies of the effectiveness of online learning for K–12 students have been published.* A systematic search of the research literature from 1994 through 2006 found no experimental or controlled quasi-experimental studies comparing the learning effects of online versus face-to-face instruction for K–12 students that provide sufficient data to compute an effect size. A subsequent search that expanded the time frame through July 2008 identified just five published studies meeting meta-analysis criteria.

The meta-analysis of 50 study effects, 43 of which were drawn from research with older learners, found that[2]

- *Students in online conditions performed modestly better, on average, than those learning the same material through traditional face-to-face instruction.* Learning outcomes for students who engaged in online learning exceeded those of students receiving face-to-face instruction, with an average effect size of +0.20 favoring online conditions.[3] The mean difference between online and face-to-face conditions across the 50 contrasts is statistically significant at the $p < .001$ level.[4] Interpretations of this result, however, should take into consideration the fact that online and face-to-face conditions generally differed on multiple dimensions, including the amount of time that learners spent on task. The advantages observed for online learning conditions therefore may be the product of aspects of those treatment conditions other than the instructional delivery medium per se.

---

[2] The meta-analysis was run also with just the 43 studies with older learners. Results were very similar to those for the meta-analysis including all 50 contrasts. Variations in findings when K-12 studies are removed are described in footnotes.

[3] The + sign indicates that the outcome for the treatment condition was larger than that for the control condition. A – sign before an effect estimate would indicate that students in the control condition had stronger outcomes than those in the treatment condition. Cohen (1992) suggests that effect sizes of .20 can be considered "small," those of approximately .50 "medium," and those of .80 or greater "large."

[4] The *p*-value represents the likelihood that an effect of this size or larger will be found by chance if the two populations under comparison do not differ. A *p*-value of less than .05 indicates that there is less than 1 chance in 20 that a difference of the observed size would be found for samples drawn from populations that do not differ.

- *Instruction combining online and face-to-face elements had a larger advantage relative to purely face-to-face instruction than did purely online instruction.* The mean effect size in studies comparing blended with face-to-face instruction was +0.35, $p < .001$. This effect size is larger than that for studies comparing purely online and purely face-to-face conditions, which had an average effect size of +0.05, $p = .46$. In fact, the learning outcomes for students in purely online conditions and those for students in purely face-to-face conditions were statistically equivalent. An important issue to keep in mind in reviewing these findings is that many studies did not attempt to equate (a) all the curriculum materials, (b) aspects of pedagogy and (c) learning time in the treatment and control conditions. Indeed, some authors asserted that it would be impossible to have done so. Hence, the observed advantage for blended learning conditions is not necessarily rooted in the media used per se and may reflect differences in content, pedagogy and learning time.

- *Effect sizes were larger for studies in which the online instruction was collaborative or instructor-directed than in those studies where online learners worked independently.*[5] The type of learning experience moderated the size of the online learning effect ($Q = 6.19$, $p < .05$).[6] The mean effect sizes for collaborative instruction (+0.25) and for instructor-directed instruction (+0.39) were significantly positive whereas the mean effect size for independent learning (+0.05) was not.

- *Most of the variations in the way in which different studies implemented online learning did not affect student learning outcomes significantly.* Analysts examined 13 online learning practices as potential sources of variation in the effectiveness of online learning compared with face-to-face instruction. Of those variables, the two mentioned above (i.e., the use of a blended rather than a purely online approach and instructor-directed or collaborative rather than independent, self-directed instruction) were the only statistically significant influences on effectiveness. The other 11 online learning practice variables that were analyzed did not affect student learning significantly. However, the relatively small number of studies contrasting learning outcomes for online and face-to-face instruction that included information about any specific aspect of implementation impeded efforts to identify online instructional practices that affect learning outcomes.

- *The effectiveness of online learning approaches appears quite broad across different content and learner types.* Online learning appeared to be an effective option for both undergraduates (mean effect of +0.30, $p < .001$) and for graduate students and professionals (+0.10, $p < .05$) in a wide range of academic and professional studies. Though positive, the mean effect size is not significant for the seven contrasts involving K–12 students, but the number of K–12 studies is too small to warrant much confidence in the mean effect estimate for this learner group. Three of the K–12 studies had

[5] Online experiences in which students explored digital artifacts and controlled the specific material they wanted to view were categorized as independent learning experiences.

[6] Online experiences in which students explored digital artifacts and controlled the specific material they wanted to view were categorized as "active" learning experiences. This contrast is not statistically significant ($p=.13$) when the five K-12 studies are removed from the analysis.

significant effects favoring a blended learning condition, one had a significant negative effect favoring face-to-face instruction, and three contrasts did not attain statistical significance. The test for learner type as a moderator variable was nonsignificant. No significant differences in effectiveness were found that related to the subject of instruction.

- *Effect sizes were larger for studies in which the online and face-to-face conditions varied in terms of curriculum materials and aspects of instructional approach in addition to the medium of instruction.* Analysts examined the characteristics of the studies in the meta-analysis to ascertain whether features of the studies' methodologies could account for obtained effects. Six methodological variables were tested as potential moderators: (a) sample size, (b) type of knowledge tested, (c) strength of study design, (d) unit of assignment to condition, (e) instructor equivalence across conditions, and (f) equivalence of curriculum and instructional approach across conditions. Only equivalence of curriculum and instruction emerged as a significant moderator variable ($Q = 6.85$, $p < .01$). Studies in which analysts judged the curriculum and instruction to be identical or almost identical in online and face-to-face conditions had smaller effects than those studies where the two conditions varied in terms of multiple aspects of instruction (+0.13 compared with +0.40, respectively). Instruction could differ in terms of the way activities were organized (for example as group work in one condition and independent work in another) or in the inclusion of instructional resources (such as a simulation or instructor lectures) in one condition but not the other.

The narrative review of experimental and quasi-experimental studies contrasting different online learning practices found that the majority of available studies suggest the following:

- *Blended and purely online learning conditions implemented within a single study generally result in similar student learning outcomes.* When a study contrasts blended and purely online conditions, student learning is usually comparable across the two conditions.

- *Elements such as video or online quizzes do not appear to influence the amount that students learn in online classes.* The research does not support the use of some frequently recommended online learning practices. Inclusion of more media in an online application does not appear to enhance learning. The practice of providing online quizzes does not seem to be more effective than other tactics such as assigning homework.

- *Online learning can be enhanced by giving learners control of their interactions with media and prompting learner reflection.* Studies indicate that manipulations that trigger learner activity or learner reflection and self-monitoring of understanding are effective when students pursue online learning as individuals.

- *Providing guidance for learning for groups of students appears less successful than does using such mechanisms with individual learners.* When groups of students are learning together online, support mechanisms such as guiding questions generally influence the way students interact, but not the amount they learn.

Conclusions

In recent experimental and quasi-experimental studies contrasting blends of online and face-to-face instruction with conventional face-to-face classes, blended instruction has been more effective, providing a rationale for the effort required to design and implement blended approaches. When used by itself, online learning appears to be as effective as conventional classroom instruction, but not more so.

However, several caveats are in order: Despite what appears to be strong support for blended learning applications, the studies in this meta-analysis do not demonstrate that online learning is superior as a *medium,* In many of the studies showing an advantage for blended learning, *the online and classroom conditions differed in terms of time spent, curriculum and pedagogy.* It was the *combination* of elements in the treatment conditions (which was likely to have included additional learning time and materials as well as additional opportunities for collaboration) that produced the observed learning advantages. At the same time, one should note that online learning is much more conducive to the expansion of learning time than is face-to-face instruction.

In addition, although the types of research designs used by the studies in the meta-analysis were strong (i.e., experimental or controlled quasi-experimental), many of the studies suffered from weaknesses such as small sample sizes; failure to report retention rates for students in the conditions being contrasted; and, in many cases, potential bias stemming from the authors' dual roles as experimenters and instructors.

Finally, the great majority of estimated effect sizes in the meta-analysis are for undergraduate and older students, not elementary or secondary learners. Although this meta-analysis did not find a significant effect by learner type, when learners' age groups are considered separately, the mean effect size is significantly positive for undergraduate and other older learners but not for K–12 students.

Another consideration is that various online learning implementation practices may have differing effectiveness for K–12 learners than they do for older students. It is certainly possible that younger online students could benefit from practices (such as embedding feedback, for example) that did not have a positive impact for college students and older learners. Without new random assignment or controlled quasi-experimental studies of the effects of online learning options for K–12 students, policy-makers will lack scientific evidence of the effectiveness of these emerging alternatives to face-to-face instruction.

1. Introduction

Online learning has roots in the tradition of distance education, which goes back at least 100 years to the early correspondence courses. With the advent of the Internet and the World Wide Web, the potential for reaching learners around the world increased greatly, and today's online learning offers rich educational resources in multiple media and the capability to support both real-time and asynchronous communication between instructors and learners as well as among different learners. Institutions of higher education and corporate training were quick to adopt online learning. Although K–12 school systems lagged behind at first, this sector's adoption of e-learning is now proceeding rapidly.

The National Center for Education Statistics estimated that 37 percent of school districts had students taking technology-supported distance education courses during school year 2004–05 (Zandberg and Lewis 2008). Enrollments in these courses (which included two-way interactive video as well as Internet-based courses), were estimated at 506,950, a 60 percent increase over the estimate based on the previous survey for 2002-03 (Selzer and Lewis 2007). Two district surveys commissioned by the Sloan Consortium (Picciano and Seaman 2007; 2008) produced estimates that 700,000 K–12 public school students took online courses in 2005–06 and over a million students did so in 2007–08—a 43 percent increase.[7] Most of these courses were at the high school level or in combination elementary-secondary schools (Zandberg and Lewis 2008).

These district numbers, however, do not fully capture the popularity of programs that are entirely online. By fall 2007, 28 states had online virtual high school programs (Tucker 2007). The largest of these, the Florida Virtual School, served over 60,000 students in 2007–08. In addition, enrollment figures for courses or high school programs that are entirely online reflect just one part of overall K–12 online learning. Increasingly, regular classroom teachers are incorporating online teaching and learning activities into their instruction.

Online learning has become popular because of its potential for providing more flexible *access to content and instruction at any time, from any place*. Frequently, the focus entails (a) increasing the availability of learning experiences for learners who cannot or choose not to attend traditional face-to-face offerings, (b) assembling and disseminating instructional content more cost-efficiently, or (c) enabling instructors to handle more students while maintaining learning outcome quality that is equivalent to that of comparable face-to-face instruction.

Different technology applications are used to support different models of online learning. One class of online learning models uses asynchronous communication tools (e.g., e-mail, threaded discussion boards, newsgroups) to allow users to contribute at their convenience. Synchronous technologies (e.g., webcasting, chat rooms, desktop audio/video technology) are used to approximate face-to-face teaching strategies such as delivering lectures and holding meetings with groups of students. Earlier online programs tended to implement one model or the other. More recent applications tend to combine multiple forms of synchronous and asynchronous online interactions as well as occasional face-to-face interactions.

---

[7] The Sloan Foundation surveys had very low response rates, suggesting the need for caution with respect to their numerical estimates.

In addition, online learning offerings are being designed to *enhance the quality of learning experiences and outcomes*. One common conjecture is that learning a complex body of knowledge effectively requires a *community of learners* (Bransford, Brown and Cocking 1999; Riel and Polin 2004; Schwen and Hara 2004; Vrasidas and Glass 2004) and that online technologies can be used to expand and support such communities. Another conjecture is that asynchronous discourse is inherently self-reflective and therefore more conducive to deep learning than is synchronous discourse (Harlen and Doubler 2004; Hiltz and Goldman 2005; Jaffee et al. 2006).

This literature review and meta-analysis have been guided by four research questions:

1. *How does the effectiveness of online learning compare with that of face-to-face instruction?*
2. *Does supplementing face-to-face instruction with online instruction enhance learning?*
3. *What practices are associated with more effective online learning?*
4. *What conditions influence the effectiveness of online learning?*

Context for the Meta-analysis and Literature Review

The meta-analysis and literature review reported here are part of the broader *Evaluation of Evidence-Based Practices in Online Learning* study that SRI International is conducting for the Policy and Program Studies Service of the U.S. Department of Education. The overall goal of the study is to provide research-based guidance to policy-makers, administrators and educators for implementing online learning for K–12 education. This literature search, analysis, and review has expanded the set of studies available for analysis by also addressing the literature concerning online learning in career technical education, medical and higher education, corporate and military training, and K–12 education.

In addition to examining the learning effects of online learning, this meta-analysis has considered the conditions and practices associated with differences in effectiveness. *Conditions* are those features of the context within which the online technology is implemented that are relatively impervious to change. Conditions include the year in which the intervention took place, the learners' demographic characteristics, the teacher's or instructor's qualifications, and state accountability systems. In contrast, *practices* concern how online learning is implemented (e.g., whether or not an online course facilitator is used). In choosing *whether or where* to use online learning (e.g., to teach mathematics for high school students, to teach a second language to elementary students), it is important to understand the degree of effectiveness of online learning under differing conditions. In deciding *how* to implement online learning, it is important to understand the practices that research suggests will increase effectiveness (e.g., community building among participants, use of an online facilitator, blending work and training).

Conceptual Framework for Online Learning

Modern online learning includes offerings that run the gamut from conventional didactic lectures or textbook-like information delivered over the Web to Internet-based collaborative role-playing in social simulations and highly interactive multiplayer strategy games. Examples include primary-grade students working on beginning reading skills over the Internet, middle school students collaborating with practicing scientists in the design and conduct of research, and teenagers who dropped out of high school taking courses online to attain the credits needed for graduation. The teachers of K–12 students may also participate in online education, logging in to online communities and reference centers and earning inservice professional development credit online.

To guide the literature search and review, the research team developed a conceptual framework identifying three key components describing online learning: (a) whether the activity served as a replacement for or an enhancement to conventional face-to-face instruction, (b) the type of learning experience (pedagogical approach), and (c) whether communication was primarily synchronous or asynchronous. Each component is described in more detail below.

One of the most basic characteristics for classifying online activities is its objective—whether the activity serves as a *replacement* for face-to-face instruction (e.g., a virtual course) or as an *enhancement* of the face-to-face learning experience (i.e., online learning activities that are part of a course given face-to-face). This distinction is important because the two types of applications have different objectives. A replacement application that is *equivalent* to conventional instruction in terms of learning outcomes is considered a success if it provides learning online without sacrificing student achievement. If student outcomes are the same whether a course is taken online or face-to-face, then online instruction can be used cost-effectively in settings where too few students are situated in a particular geographic locale to warrant an on-site instructor (e.g., rural students, students in specialized courses). In contrast, online enhancement activities that produce learning outcomes that are only equivalent to (not better than) those resulting from face-to-face instruction alone would be considered a waste of time and money because the addition does not improve student outcomes.

A second important dimension is the type of learning *experience*, which depends on who (or what) determines the way learners acquire knowledge. Learning experiences can be classified in terms of the amount of control that the student has over the content and nature of the learning activity. In traditional didactic or *expository learning* experiences, content is transmitted to the student by a lecture, written material, or other mechanisms. Such conventional instruction is often contrasted with *active learning* in which the student has control of what and how he or she learns. Another category of learning experiences stresses collaborative or *interactive learning* activity in which the nature of the learning content is emergent as learners interact with one another and with a teacher or other knowledge sources. Technologies can support any of these three types of learning experience:

- Expository instruction—Digital devices transmit knowledge.
- Active learning—The learner builds knowledge through inquiry-based manipulation of digital artifacts such as online drills, simulations, games, or microworlds.

- Interactive learning—The learner builds knowledge through inquiry-based collaborative interaction with other learners; teachers become co-learners and act as facilitators.

This dimension of learning-experience type is closely linked to the concept of learner control explored by Zhang (2005). Typically, in expository instruction, the technology delivers the content. In active learning, the technology allows students to control digital artifacts to explore information or address problems. In interactive learning, technology mediates human interaction either synchronously or asynchronously; learning emerges through interactions with other students and the technology.

The learner-control category of interactive learning experiences is related to the so-called "fifth generation" of distance learning, which stresses a flexible combination of independent and group learning activities. Researchers are now using terms such as "distributed learning" (Dede 2006) or "learning communities" to refer to orchestrated mixtures of face-to-face and virtual interactions among a cohort of learners led by one or more instructors, facilitators or coaches over an extended period of time (from weeks to years).

Finally, a third characteristic commonly used to categorize online learning activities is the extent to which the activity is *synchronous*, with instruction occurring in real time whether in a physical or a virtual place, or *asynchronous*, with a time lag between the presentation of instructional stimuli and student responses. Exhibit 1 illustrates the three dimensions in the framework guiding this meta-analysis of online learning offerings. The descriptive columns in the table illustrate uses of online learning comprising dimensions of each possible combination of the learning experience, synchronicity, and objective (an alternative or a supplement to face-to-face instruction).

Exhibit 1. Conceptual Framework for Online Learning

Learning Experience Dimension	Synchronicity	Face-to-Face Alternative	Face-to-Face Enhancement
Expository	Synchronous	Live, one-way webcast of online lecture course with limited learner control (e.g., students proceed through materials in set sequence)	Viewing webcasts to supplement in-class learning activities
	Asynchronous	Math course taught through online video lectures that students can access on their own schedule	Online lectures on advanced topics made available as a resource for students in a conventional math class
Active	Synchronous	Learning how to troubleshoot a new type of computer system by consulting experts through live chat	Chatting with experts as the culminating activity for a curriculum unit on network administration
	Asynchronous	Social studies course taught entirely through Web quests that explore issues in U.S. history	Web quest options offered as an enrichment activity for students completing their regular social studies assignments early
Interactive	Synchronous	Health-care course taught entirely through an online, collaborative patient management simulation that multiple students interact with at the same time	Supplementing a lecture-based course through a session spent with a collaborative online simulation used by small groups of students
	Asynchronous	Professional development for science teachers through “threaded” discussions and message boards on topics identified by participants	Supplemental, threaded discussions for pre-service teachers participating in a face-to-face course on science methods

Exhibit reads: Online learning applications can be characterized in terms of (a) the kind of learning experience they provide, (b) whether computer-mediated instruction is primarily synchronous or asynchronous and (c) whether they are intended as an alternative or a supplement to face-to-face instruction.

Many other features also apply to online learning, including the type of setting (classroom, home, informal), the nature of the content (both the subject area and the type of learning such as fact, concept, procedure or strategy), and the technology involved (e.g., audio/video streaming, Internet telephony, podcasting, chat, simulations, videoconferencing, shared graphical whiteboard, screen sharing).

The dimensions in the framework in Exhibit 1 were derived from prior meta-analyses in distance learning. Bernard et al. (2004) found advantages for asynchronous over synchronous distance education. In examining a different set of studies, Zhao et al. (2005) found that studies of distance-learning applications that combined synchronous and asynchronous communication tended to report more positive effects than did studies of distance learning applications with just one of these interaction types.[8] Zhao et al. also found (a) advantages for blended learning (called "Face-to-Face Enhancement" in the Exhibit 1 framework) over purely online learning experiences and (b) advantages for courses with more instructor involvement compared with more "canned" applications that provide expository learning experiences. Thus, the three dimensions in Exhibit 1 capture some of the most important kinds of variation in distance learning and together provide a manageable framework for differentiating among the broad array of online activities in practice today.

Findings From Prior Meta-Analyses

Prior meta-analyses of distance education (including online learning studies and studies of other forms of distance education) and of Web-based or online learning have been conducted. Overall, results from Bernard et al. (2004) and other reviews of the distance education literature (Cavanaugh 2001; Moore 1994) indicate no significant differences in effectiveness between distance education and face-to-face education, suggesting that distance education, when it is the only option available, can successfully replace face-to-face instruction. Findings of a recent meta-analysis of job-related courses comparing Web-based and classroom-based learning (Sitzmann et al. 2006) were even more positive. They found online learning to be superior to classroom-based instruction in terms of declarative knowledge outcomes, with the two being equivalent in terms of procedural learning.

However, a general conclusion that distance and face-to-face instruction result in essentially similar learning ignores differences in findings across various studies. Bernard et al. (2004) found tremendous variability in effect sizes (an effect size is the difference between the mean for the treatment group and the mean for the control group, divided by the pooled standard deviation), which ranged from –1.31 to +1.41.[9] From their meta-analysis, which included coding for a wide range of instructional and other characteristics, the researchers concluded that selected

[8] Both of these meta-analyses included video-based distance learning as well as Web-based learning and also included studies in which the outcome measure was student satisfaction, attitude or other nonlearning measures. The meta-analysis reported here is restricted to an analysis of effect sizes for objective student learning measures in experimental, controlled quasi-experimental, and crossover studies of applications with Web-based components.

[9] Cohen (1992) suggests that effect sizes of .20 can be considered "small," those of approximately .50 "medium," and those of .80 or greater "large."

conditions and practices were associated with differences in outcomes. For example, they found that distance education that used synchronous instruction was significantly negative in its effect, with an average effect size of –0.10, whereas the average effect size for studies using asynchronous instruction was significantly positive (+0.05). However, the studies that Bernard et al. categorized as using synchronous communication involved "yoked" classrooms; that is, the instructor's classroom was the center of the activity, and one or more distant classrooms interacted with it in "hub and spoke" fashion. These satellite classes are markedly different from today's Web-based communication among the multiple nodes in a "learning network."

Machtmes and Asher's earlier (2000) meta-analysis of telecourses sheds light on this issue.[10] Although detecting no difference between distance and face-to-face learning overall, they found results more favorable for telecourses when classrooms had two-way, as opposed to one-way, interactions.

Although earlier meta-analyses of distance education found it equivalent to classroom instruction (as noted above), several reviewers have suggested that this pattern may change. They argue that online learning as practiced in the 21st century can be expected to outperform earlier forms of distance education in terms of effects on learning (Zhao et al. 2005).

The meta-analysis reported here differs from earlier meta-analyses because its focus has been restricted to studies that did the following:

- Investigated significant use of the Web for instruction
- Had an objective learning measure as the outcome measure
- Met higher quality criteria in terms of study design (i.e., an experimental or controlled quasi-experimental design)

Structure of the Report

Chapter 2 describes the methods used in searching for appropriate research articles, in screening those articles for relevance and study quality, in coding study features, and in calculating effect sizes. Chapter 3 describes the 50 study effects identified through the article search and screening and presents findings in the form of effect sizes for studies contrasting purely online or blended learning conditions with face-to-face instruction. Chapter 4 provides a qualitative narrative synthesis of research studies comparing variations of online learning interventions. Finally, chapter 5 discusses the implications of the literature search and meta-analysis for future studies of online learning and for K–12 online learning practices.

[10] Like the present meta-analysis, Machtmes and Asher limited their study corpus to experiments or quasi-experiments with an achievement measure as the learning outcome.

2. Methodology

This chapter describes the procedures used to search for, screen and code controlled studies of the effectiveness of online learning. The products of these search, screening and coding activities were used for the meta-analysis and narrative literature review, which are described in chapters 3 and 4, respectively.

Definition of Online Learning

For this review, *online learning* is defined as learning that takes place partially or entirely over the Internet. This definition excludes purely print-based correspondence education, broadcast television or radio, videoconferencing, videocassettes, and stand-alone educational software programs that do not have a significant Internet-based instructional component.

In contrast to previous meta-analyses, this review distinguishes between two purposes for online learning:

- Learning conducted totally online as a substitute or *alternative* to face-to-face learning
- Online learning components that are combined or *blended* (sometimes called "hybrid") with face-to-face instruction to provide learning *enhancement*

As indicated in chapter 1, this distinction was made because of the different implications of finding a null effect (i.e., no difference in effects between the treatment and the control group) under the two circumstances. Equivalence between online learning and face-to-face learning justifies using online alternatives, but online enhancements need to be justified by superior learning outcomes. These two purposes of online learning defined the first two categories of study in the literature search:

- Studies comparing an online learning condition with a face-to-face control condition (Category 1)
- Studies comparing a blended condition with a face-to-face control condition without the online learning components (Category 2).

In addition, researchers sought experimental and controlled quasi-experimental studies that compared the effectiveness of different online learning practices. This third study category consisted of the following:

- Studies testing the learning effects of variations in online learning practices such as online learning with and without interactive video (Category 3).

Data Sources and Search Strategies

Relevant studies were located through a comprehensive search of publicly available literature published from 1996 through July 2008.[11] Searches of dissertations were limited to those published from 2005 through July 2008 to allow researchers to use UMI ProQuest Digital Dissertations for retrieval.

Electronic Database Searches

Using a common set of keywords, searches were performed in five electronic research databases: *ERIC, PsycINFO, PubMed, ABI/INFORM,* and *UMI ProQuest Digital Dissertations*. The appendix lists the terms used for the initial electronic database search and for additional searches for studies of online learning in the areas of career technical education and teacher education.

Additional Search Activities

The electronic database searches were supplemented with a review of articles cited in recent meta-analyses and narrative syntheses of research on distance learning (Bernard et al. 2004; Cavanaugh et al. 2004; Childs 2001; Sitzmann et al. 2006; Tallent-Runnels et al. 2006; WestEd with Edvance Research 2008; Wisher and Olson 2003; Zhao et al. 2005), including those for teacher professional development and career technical education (Whitehouse et al. 2006; Zirkle 2003). The analysts examined references from these reviews to identify studies that might meet the criteria for inclusion in the present review.

Abstracts were manually reviewed for articles published since 2005 in the following key journals: *American Journal of Distance Education, Journal of Distance Education* (Canada), *Distance Education* (Australia), *International Review of Research in Distance and Open Education,* and *Journal of Asynchronous Learning Networks.* In addition, the *Journal of Technology and Teacher Education* and *Career and Technical Education Research* were searched manually. Finally, the Google Scholar search engine was used with a series of keywords related to online learning (available from the authors). Article abstracts retrieved through these additional search activities were reviewed to remove duplicates of articles identified earlier.

[11] Literature searches were performed in two waves: in March 2007 for studies published from 1996–2006 and in July 2008 for studies published from 2007 to July 2008.

Screening Process

Screening of the research studies obtained through the search process described above was carried out in two stages. The intent of the two-stage approach was to gain efficiency without risking exclusion of potentially relevant, high-quality studies of online learning effects.

Initial Screen for Abstracts From Electronic Databases

The initial electronic database searches (excluding the additional searches conducted for teacher professional development and career technical education) yielded 1,132 articles.[12] Citation information and abstracts of these studies were examined to ascertain whether they met the following three initial inclusion criteria:

1. *Does the study address online learning as this review defines it?*
2. *Does the study appear to use a controlled design (experimental/quasi-experimental design)?*
3. *Does the study report data on student achievement or another learning outcome?*

At this early stage, analysts gave studies "the benefit of the doubt," retaining those that were not clearly outside the inclusion criteria on the basis of their citations and abstracts. As a result of this screening, 316 articles were retained and 816 articles were excluded. During this initial screen, 45 percent of the excluded articles were removed primarily because they did not have a controlled design. Twenty-six percent of excluded articles were eliminated because they did not report learning outcomes for treatment and control groups. Twenty-three percent were eliminated because their intervention did not qualify as online learning, given the definition used for this meta-analysis and review. The remaining six percent of excluded articles posed other difficulties, such as being written in a language other than English.

Full-text Screen

From the other data sources (i.e., references in earlier reviews, manual review of key journals, recommendation from a study advisor, and Google Scholar searches), researchers identified and retrieved an additional 186 articles, yielding a total of 502 articles that they subjected to a full-text screening for possible inclusion in the analysis. Nine analysts who were trained on a set of full-text screening criteria reviewed the 502 articles for both topical relevance and study quality.

A study had to meet *content relevance* criteria to be included in the meta-analysis. Thus, qualifying studies had to

1. *Involve learning that took place over the Internet*. The use of the Internet had to be a major part of the intervention. Studies in which the Internet was only an incidental component of the intervention were excluded. In operational terms, to qualify as *online learning*, a study treatment needed to provide at least a quarter of the

[12] This number includes multiple instances of the same study identified in different databases.

instruction/learning of the content assessed by the study's learning measure by means of the Internet.

2. *Contrast conditions that varied in terms of use of online learning*. Learning outcomes had to be compared against conditions falling into at least one of two study categories: Category 1, online learning compared with offline/face-to-face learning, and Category 2, a combination of online plus offline/face-to-face learning (i.e., blended learning) compared with offline/face-to-face learning alone.
3. *Describe an intervention study that had been completed*. Descriptions of study designs, evaluation plans or theoretical frameworks were excluded. The length of the intervention/treatment could vary from a few hours to a quarter, semester, year or longer.
4. *Report a learning outcome that was measured for both treatment and control groups*. A learning outcome needed to be measured in the same way across study conditions. A study was excluded if it explicitly indicated that different examinations were used for the treatment and control groups. The measure had to be objective and direct; learner or teacher/instructor self-report of learning was not considered a direct measure. Examples of learning outcome measures included scores on standardized tests, scores on researcher-created assessments, grades/scores on teacher-created assessments (e.g., assignments, midterm/final exams), and grades or grade point averages. Examples of learning outcome measures for teacher learners (in addition to those accepted as student outcomes) included assessments of content knowledge, analysis of lesson plans or other materials related to the intervention, observation (or logs) of class activities, analysis of portfolios, or supervisor's rating of job performance. Studies that used only nonlearning outcome measures (e.g., attitude, retention, attendance, level of learner/instructor satisfaction) were excluded.

Studies also had to meet basic *Quality* (method) criteria to be included. Thus, qualifying studies had to

5. *Use a controlled design* (experimental or quasi-experimental). Design studies, exploratory studies or case studies that did not use a controlled research design were excluded. For quasi-experimental designs, the analysis of the effects of the intervention had to include statistical controls for possible differences between the treatment and control groups in terms of prior achievement.
6. *Report sufficient data for effect size calculation or estimation* as specified in the guidelines provided by the What Works Clearinghouse (2007) and by Lipsey and Wilson (2000).

Studies that contrasted different versions of online learning (Category 3) needed to meet Criteria 1 and 3–5 to be included in the narrative research summary.

An analyst read each full text, and all borderline cases were discussed and resolved either at project meetings or through consultation with task leaders. To prevent studies from being mistakenly screened out, two analysts coded studies on features that were deemed to require significant degrees of inference. These features consisted of the following:

- Failure to have students use the Internet for a significant portion of the time that they spent learning the content assessed by the study's learning measure
- Lack of statistical control for prior abilities in quasi-experiments

From the 502 articles, analysts identified 522 independent studies (some articles reported more than one study). When the same study was reported in different publication formats (e.g., conference paper and journal article), only the more formal journal article was retained for the analysis.

Of the 522 studies, 176 met all the criteria of the full-text screening process. Exhibit 2 shows the bases for exclusion for the 346 studies that did not meet all the criteria.

Exhibit 2. Bases for Excluding Studies During the Full-Text Screening Process

Primary Reason for Exclusion	Number Excluded	Percentage Excluded
Did not use statistical control	137	39
Was not online as defined in this review	90	26
Did not analyze learning outcomes	52	15
Did not have a comparison group that received a comparable treatment	22	7
Did not fit into any of the three study categories	39	11
Excluded for other reasons[a]	6	2

Exhibit reads: The most common reason for a study's exclusion from the analysis was failure to use statistical control (in a quasi-experiment).
[a]Other reasons for exclusion included (a) did not provide enough information, (b) was written in a language other than English, and (c) used different learning outcome measures for the treatment and control groups.

Effect Size Extraction

Of the 176 independent studies, 99 had at least one contrast between online learning and face-to-face/offline learning (Category 1) or between blended learning and face-to-face/offline learning (Category 2). These studies were subjected to quantitative analysis to extract effect sizes.

Of the 99 studies, only nine were conducted with K–12 students (Chang 2008; Englert et al. 2007; Long and Jennings 2005; O'Dwyer, Carey and Kleiman 2007; Parker 1999; Rockman et al. 2007; Stevens 1999; Sun, Lin and Yu 2008; Uzunboylu 2004). Of them, four were excluded from the meta-analysis: Chang (2008), Parker (1999), and Uzunboylu (2004) did not provide sufficient statistical data to compute effect sizes, and the Stevens (1999) study was a quasi-experiment without a statistical control for potential existing differences in achievement.

An effect size is similar to a *z*-score in that it is expressed in terms of units of standard deviation. It is defined as the difference between the treatment and control means, divided by the pooled standard deviation. Effect sizes can be calculated (a) from the means and standard deviations for the two groups or (b) on the basis of information provided in statistical tests such as *t*-tests and analyses of variance. Following the guidelines from the What Works Clearinghouse (2007) and Lipsey and Wilson (2000), numerical and statistical data contained in the studies were extracted so that Comprehensive Meta-Analysis software (Biostat Solutions 2006) could be used to calculate effect sizes (*g*). The precision of each effect estimate was determined by using the estimated standard error of the mean to calculate the 95-percent confidence interval for each effect.

The review of the 99 studies to obtain the data for calculating effect size produced 50 independent effect sizes (27 for Category 1 and 23 for Category 2) from 45 studies. Fifty-four studies did not report sufficient data to support calculating effect size.

Coding of Study Features

All studies that provided enough data to compute an effect size were coded for their study features and for study quality. Building on the project's conceptual framework and the coding schemes used in several earlier meta-analyses (Bernard et al. 2004; Sitzmann et al. 2006), a coding structure was developed and pilot-tested with several studies. The top-level coding structure, incorporating refinements made after pilot testing, is shown in Exhibit A-4 of the appendix.

To determine interrater reliability, two researchers coded 20 percent of the studies, achieving an interrater reliability of 86 percent across those studies. Analysis of coder disagreements resulted in the refinement of some definitions and decision rules for some codes; other codes that required information that articles did not provide or that proved difficult to code reliably were eliminated (e.g., whether or not the instructor was certified). A single researcher coded the remaining studies.

Data Analysis

Before combining effects from multiple contrasts, effect sizes were weighted to avoid undue influence of studies with small sample sizes (Hedges and Olkin 1985). For the total set of 50 contrasts and for each subset of contrasts being investigated, a weighted mean effect size (Hedges' *g*+) was computed by weighting the effect size for each study contrast by the inverse of its variance. The precision of each mean effect estimate was determined by using the estimated standard error of the mean to calculate the 95 percent confidence interval. Using a fixed-effects model, the heterogeneity of the effect size distribution (the *Q*-statistic) was computed to indicate the extent to which variation in effect sizes was not explained by sampling error alone.

Next, a series of post-hoc subgroup and moderator variable analyses were conducted using the Comprehensive Meta-Analysis software. A mixed-effects model was used for these analyses to model within-group variation.[13] A between-group heterogeneity statistic (Q_{Between}) was computed to test for statistical differences in the weighted mean effect sizes for various subsets of the effects (e.g., studies using blended as opposed to purely online learning for the treatment group). Chapter 3 describes the results of these analyses.

[13] Meta-analysts need to choose between a mixed-effects and a fixed-effects model for investigating moderator variables. A fixed-effects analysis is more sensitive to differences related to moderator variables, but has a greater likelihood of producing Type I errors (falsely rejecting the null hypothesis). The mixed-effects model reduces the likelihood of Type I errors by adding a random constant to the standard errors, but does so at the cost of increasing the likelihood of Type II errors (incorrectly accepting the null hypothesis). Analysts chose the more conservative mixed-effects model for this investigation of moderator variables.

3. Findings

This chapter presents the results of the meta-analysis of controlled studies that compared the effectiveness of online learning with that of face-to-face instruction. The next chapter presents a narrative synthesis of studies that compared different versions of online learning with each other rather than with a face-to-face control condition.

Nature of the Studies in the Meta-Analysis

As indicated in chapter 2, 50 independent effect sizes could be abstracted from the study corpus of 45 studies.[14] The number of students in the studies included in the meta-analysis ranged from 16 to 1,857, but most of the studies were modest in scope. Although large-scale applications of online learning have emerged, only five studies in the meta-analysis corpus included more than 400 learners. The types of learners in these studies were about evenly split between students in college or earlier years of education and learners in graduate programs or professional training. The average learner age ranged from 13 to 44. Nearly all the studies involved formal instruction, with the most common subject matter being medicine or health care. Other content types included computer science, teacher education, social science, mathematics, languages, science and business. Roughly half of the learners were taking the instruction for credit or as an academic requirement. Of the 48 contrasts for which the study indicated the length of instruction, 19 involved instructional time frames of less than a month and the remainder involved longer periods.

In terms of instructional features, the online learning conditions in these studies were less likely to be instructor-directed (8 contrasts) than they were to be student-directed, independent learning (17 contrasts) or interactive and collaborative in nature (22 contrasts). Online learners typically had opportunities to practice skills or test their knowledge (41 effects were from studies reporting such opportunities). Opportunities for learners to receive feedback were less common; however, it was reported in the studies associated with 23 effects. The opportunity for online learners to have face-to-face contact with the instructor during the time frame of the course was present in the case of 21 out of 50 effects. The details of instructional media and communication options available to online learners were absent in many of the study narratives. Among the 50 contrasts, analysts could document the presence of one-way video or audio in the online condition for 14 effects. Similarly, 16 contrasts involved online conditions which allowed students to communicate with the instructor with asynchronous communication only; 8 allowed both asynchronous and synchronous online communication; and 26 contrasts came from studies that did not document the types of online communication provided between the instructor and learners.

[14] After the first literature search, which yielded 29 independent effects, the research team ran additional analyses to find out how many more studies could be included if the study design criterion were relaxed to include quasi-experiments with pre- and posttests with no statistical adjustments made for preexisting differences. The relaxed standard would have increased the corpus for analysis by just 10 studies, nearly all of which were in Category 1 and which had more positive effect sizes than the Category 1 studies with stronger analytic designs. Analysts decided not to include those studies in the meta-analysis. Instead, the study corpus was enlarged by conducting a second literature search in July 2008.

Among the 50 individual contrasts between online and face-to-face instruction, 11 were significantly positive, favoring the online or blended learning condition. Three significant negative effects favored traditional face-to-face instruction. The fact that multiple comparisons were conducted should be kept in mind when interpreting this pattern of findings. Because analysts used a $\alpha < .05$ level of significance for testing differences, one would expect approximately 1 in 20 contrasts to show a significant difference by chance alone. For 50 contrasts, then, one would expect 2 or 3 significant differences by chance. The finding of 3 significant contrasts favoring face-to-face instruction is within the range one would expect by chance; the 11 contrasts favoring online or hybrid instruction exceeds what one would expect by chance.

Exhibit 3 illustrates the 50 effect sizes derived from the 45 articles.[15] Exhibits 4a and 4b present the effect sizes for Category 1 (purely online versus face-to-face) and Category 2 (blended versus face-to-face) studies, respectively, along with standard errors, statistical significance, and the 95-percent confidence interval.

Main Effects

The overall finding of the meta-analysis is that classes with online learning (whether taught completely online or blended) on average produce stronger student learning outcomes than do classes with solely face-to-face instruction. The mean effect size for all 50 contrasts was +0.20, $p < .001$.

The conceptual framework for this study, which distinguishes between purely online and blended forms of instruction, calls for creating subsets of the effect estimates to address two more nuanced research questions:

- *How does the effectiveness of online learning compare with that of face to-face instruction?* Looking only at the 27 Category 1 effects that compared a purely online condition with face-to-face instruction, analysts found a mean effect of +0.05, $p = .46$. This finding is similar to that of previous summaries of distance learning (generally from pre-Internet studies), in finding that instruction conducted entirely online is as effective as classroom instruction but no better.

[15] Some references appear twice in Exhibit 3 because multiple effect sizes were extracted from the same article. Davis et al. (1999) and Caldwell (2006) each included two contrasts—online versus face-to-face (Category 1) and blended versus face-to-face (Category 2). Rockman et al. (2007) and Schilling et al. (2006) report findings for two distinct learning measures. Long and Jennings (2005) report findings from two distinct experiments, a "wave 1" in which teachers were implementing online learning for the first time and a "wave 2" in which teachers implemented online learning a second time with new groups of students.

- *Does supplementing face-to-face instruction with online instruction enhance learning?* For the 23 Category 2 contrasts that compared blended conditions of online plus face-to-face learning with face-to-face instruction alone, the mean effect size of +0.35 was significant ($p < .0001$). Blends of online and face-to-face instruction, on average, had stronger learning outcomes than did face-to-face instruction alone.

A test of the difference between Category 1 and Category 2 studies found that the mean effect size was larger for contrasts pitting blended learning against face-to-face instruction ($g+ = +0.35$) than for those of purely online versus face-to-face instruction ($g+ = +0.05$); the difference between the two subsets of studies was statistically significant ($Q = 8.37$, $p < .01$).

Exhibit 3. Effect Sizes for Contrasts in the Meta-Analysis

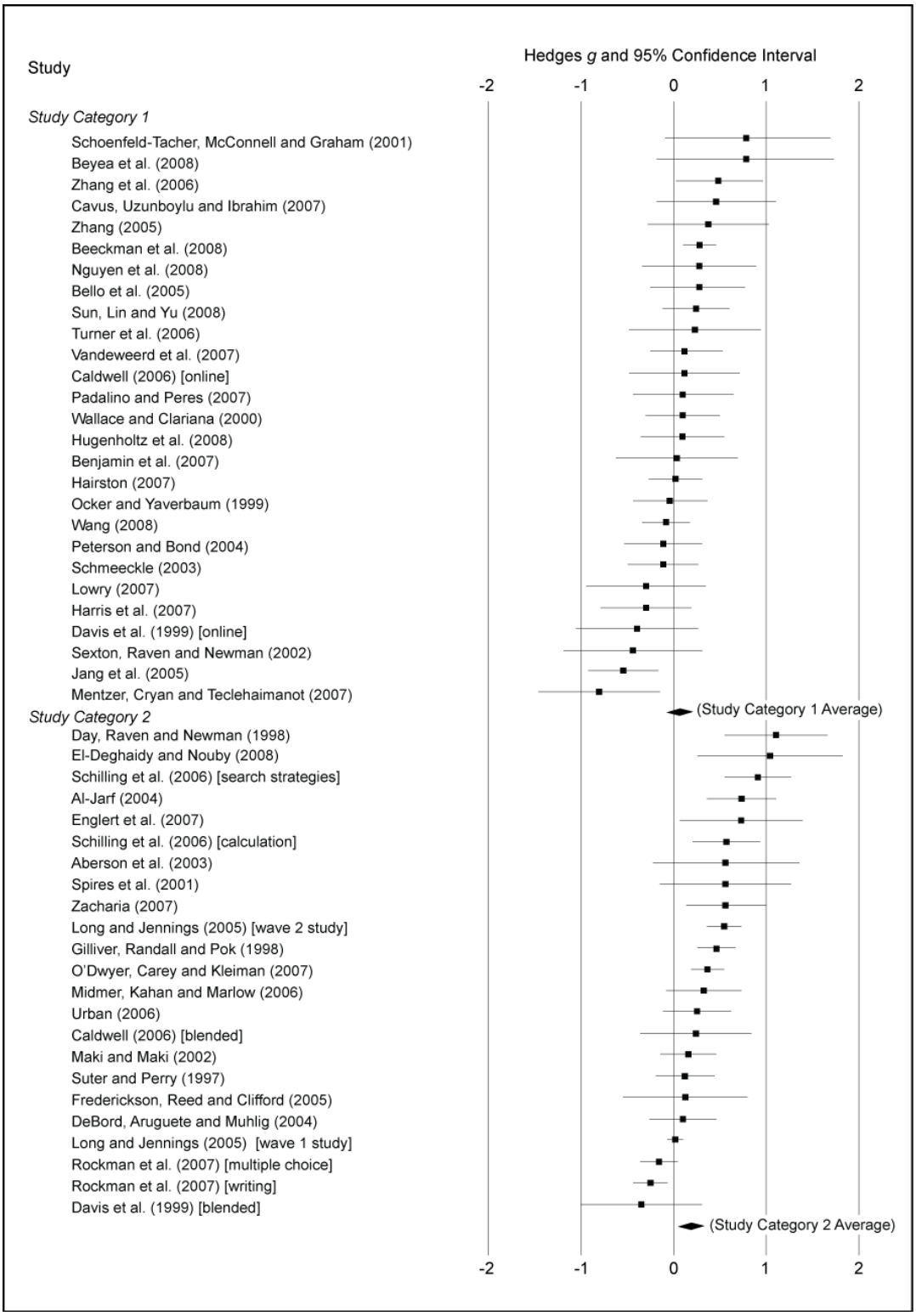

Exhibit reads: The effect size estimate for Schoenfeld-Tacher, McConnell and Graham (2001) was +0.80 with a 95 percent probability that the true effect size lies between -0.10 and +1.70.

Exhibit 4a. Purely Online Versus Face-to-Face (Category 1) Studies Included in the Meta-Analysis

Authors	Title	Effect Size		95-Percent Confidence Interval		Test of Null Hypothesis (2-tail)	Retention Rate (percentage)		Number of Units Assigned[a]
		g	*SE*	Lower Limit	Upper Limit	*Z*-Value	Online	Face-to-Face	
Beeckman et al. (2008)	Pressure ulcers: E-learning to improve classification by nurses and nursing students	+0.294	0.097	0.104	0.484	3.03**	Unknown	Unknown	426 participants
Bello et al. (2005)	Online vs. live methods for teaching difficult airway management to anesthesiology residents	+0.278	0.265	-0.241	0.797	1.05	100	100	56 participants
Benjamin et al. (2007)	A randomized controlled trial comparing Web to in-person training for child care health consultants	+0.046	0.340	-0.620	0.713	0.14	Unknown	Unknown	23 participants
Beyea et al. (2008)	Evaluation of a particle repositioning maneuver Web-based teaching module	+0.790	0.493	-0.176	1.756	1.60	Unknown	Unknown	17–20 participants[b]
Caldwell (2006)	A comparative study of traditional, Web-based and online instructional modalities in a computer programming course	+0.132	0.310	-0.476	0.740	0.43	100	100	60 students
Cavus, Uzonboylu and Ibrahim (2007)	Assessing the success rate of students using a learning management system together with a collaborative tool in Web-based teaching of programming languages	+0.466	0.335	-0.190	1.122	1.39	Unknown	Unknown	54 students
Davis et al. (1999)	Developing online courses: A comparison of Web-based instruction with traditional instruction	-0.379	0.339	-1.042	0.285	-1.12	Unknown	Unknown	2 courses/ classrooms
Hairston (2007)	Employees' attitudes toward e-learning: Implications for policy in industry environments	+0.028	0.155	-0.275	0.331	0.18	70	58.33	168 participants
Harris et al. (2008)	Educating generalist physicians about chronic pain with live experts and online education	-0.285	0.252	-0.779	0.209	-1.13	84.21	94.44	62 participants
Hugenholtzet al. (2008)	Effectiveness of e-learning in continuing medical education for occupational physicians	+0.106	0.233	-0.351	0.564	0.46	Unknown	Unknown	72 participants
Jang et al. (2005)	Effects of a Web-based teaching method on undergraduate nursing students' learning of electrocardiography	-0.530	0.197	-0.917	-0.143	-2.69**	85.71	87.93	105 students

Exhibit 4a. Purely Online Versus Face-to-Face (Category 1) Studies Included in the Meta-Analysis (continued)

Authors	Title	Effect Size		95-Percent Confidence Interval		Test of Null Hypothesis (2-tail)	Retention Rate (percentage)		Number of Units Assigned[a]
		g	*SE*	Lower Limit	Upper Limit	*Z*-Value	Online	Face-to-Face	
Lowry (2007)	Effects of online versus face-to-face professional development with a team-based learning community approach on teachers' application of a new instructional practice	-0.281	0.335	-0.937	0.370	-0.84	80	93.55	53 students
Mentzer, Cryan and Teclehaimanot (2007)	A comparison of face-to-face and Web-based classrooms	-0.796	0.339	-1.460	-0.131	-2.35*	Unknown	Unknown	36 students
Nguyen et al. (2008)	Randomized controlled trial of an Internet-based versus face-to-face dyspnea self-management program for patients with chronic obstructive pulmonary disease: Pilot study	+0.292	0.316	-0.327	0.910	0.93	Unknown	Unknown	39 participants
Ocker and Yaverbaum (1999)	Asynchronous computer-mediated communication versus face-to-face collaboration: Results on student learning, quality and satisfaction	-0.030	0.214	-0.449	0.389	-0.14	Unknown	Unknown	43 students
Padalino and Peres (2007)	E-learning: A comparative study for knowledge apprehension among nurses	0.115	0.281	-0.437	0.666	0.41	Unknown	Unknown	49 participants
Peterson and Bond (2004)	Online compared to face-to-face teacher preparation for learning standards-based planning skills	-0.100	0.214	-0.520	0.320	-0.47	Unknown	Unknown	4 sections
Schmeeckle (2003)	Online training: An evaluation of the effectiveness and efficiency of training law enforcement personnel over the Internet	-0.106	0.198	-0.494	0.282	-0.53	Unknown	Unknown	101 students
Schoenfeld-Tacher, McConnell and Graham (2001)	Do no harm: A comparison of the effects of online vs. traditional delivery media on a science course	+0.800	0.459	-0.100	1.700	1.74	100	99.94	Unknown

Exhibit 4a: Purely Online versus Face-to-Face (Category 1) Studies Included in the Meta-analysis (continued)

Authors	Title	Effect Size		95-Percent Confidence Interval		Test of Null Hypothesis (2-tail)	Retention Rate (percentage)		Number of Units Assigned[a]
		g	*SE*	Lower Limit	Upper Limit	*Z*-Value	Online	Face-to-Face	
Sexton, Raven and Newman (2002)	A comparison of traditional and World Wide Web methodologies, computer anxiety, and higher order thinking skills in the inservice training of Mississippi 4-H extension agents	-0.422	0.385	-1.177	0.332	-1.10	Unknown	Unknown	26 students
Sun, Lin and Yu (2008)	A study on learning effect among different learning styles in a Web-based lab of science for elementary school students	+0.260	0.188	-0.108	0.628	1.38	Unknown	Unknown	4 classrooms
Turner et al. (2006)	Web-based learning versus standardized patients for teaching clinical diagnosis: A randomized, controlled, crossover trial	+0.242	0.367	-0.477	0.960	0.66	Unknown	Unknown	30 students
Vandeweerd et al. (2007)	Teaching veterinary radiography by e-learning versus structured tutorial: A randomized, single-blinded controlled trial	+0.144	0.207	-0.262	0.550	0.70	Unknown	Unknown	92 students
Wallace and Clariana (2000)	Achievement predictors for a computer-applications module delivered online	+0.109	0.206	-0.295	0.513	0.53	Unknown	Unknown	4 sections
Wang (2008)	Developing and evaluating an interactive multimedia instructional tool: Learning outcomes and user experiences of optometry students	-0.071	0.136	-0.338	0.195	-0.53	Unknown	Unknown	4 sections[c]
Zhang (2005)	Interactive multimedia-based e-learning: A study of effectiveness	+0.381	0.339	-0.283	1.045	1.12	Unknown	Unknown	51 students
Zhang et al. (2006)	Instructional video in e-learning: Assessing the effect of interactive video on learning effectiveness	+0.498	0.244	0.020	0.975	2.04*	Unknown	Unknown	69 students

Exhibit reads: The effect size for the Hugenholtz et al. (2008) study of online medical education was +0.11, which was not significantly different from 0.

*p < .05, ** p < .01, SE = Standard error

[a] The number given represents the assigned units at study conclusion. It excludes units that attrited.

[b] Two outcome measures were used to compute one effect size. The first outcome measure was completed by 17 participants, and the second outcome measure was completed by 20 participants.

[c] This study is a crossover study. The number of units represents those assigned to treatment and control conditions in the first round.

Exhibit 4b. Blended Versus Face-to-Face (Category 2) Studies Included in the Meta-Analysis

Authors	Title	Effect Size		95-Percent Confidence Interval		Test of Null Hypothesis (2-tail)	Retention Rate (percentage)		Number of Units Assigned[a]
		g	*SE*	Lower Limit	Upper Limit	*Z*-Value	Online	Face-to-Face	
Aberson et al. (2003)	Evaluation of an interactive tutorial for teaching hypothesis testing concepts	+0.580	0.404	-0.212	1.372	1.44	Unknown	.75	2 sections
Al-Jarf (2004)	The effects of Web-based learning on struggling EFL college writers	+0.740	0.194	0.360	1.120	3.82***	Unknown	Unknown	113 students
Caldwell (2006)	A comparative study of traditional, Web-based and online instructional modalities in a computer programming course	+0.251	0.311	-0.359	0.861	0.81	100	100	60 students
Davis et al. (1999)	Developing online courses: A comparison of Web-based instruction with traditional instruction	-0.335	0.338	-0.997	0.327	-0.99	Unknown	Unknown	2 courses/ classrooms
Day, Raven and Newman (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	+1.113	0.289	0.546	1.679	3.85***	89.66	96.55	2 sections
DeBord, Aruguete and Muhlig (2004)	Are computer-assisted teaching methods effective?	+0.110	0.188	-0.259	0.479	0.69	Unknown	Unknown	112 students
El-Deghaidy and Nouby (2008)	Effectiveness of a blended e-learning cooperative approach in an Egyptian teacher education program	+1.049	0.406	0.253	1.845	2.58**	Unknown	Unknown	26 students
Englert et al. (2007)	Scaffolding the writing of students with disabilities through procedural facilitation using an Internet-based technology	+0.740	0.345	0.064	1.416	2.15*	Unknown	Unknown	6 classrooms from 5 urban schools
Frederickson, Reed and Clifford (2005)	Evaluating Web-supported learning versus lecture-based teaching: Quantitative and qualitative perspectives	+0.138	0.345	-0.539	0.814	0.40	Unknown	Unknown	2 sections
Gilliver, Randall and Pok (1998)	Learning in cyberspace: Shaping the future	+0.477	0.111	0.260	0.693	4.31***	Unknown	Unknown	24 classes
Long and Jennings (2005) [Wave 1][c]	The effect of technology and professional development on student achievement	+0.025	0.046	-0.066	0.116	0.53	Unknown	Unknown	9 schools

Exhibit 4b: Blended versus Face-to-Face (Category 2) Studies Included in the Meta-analysis (continued)

Authors	Title	Effect Size		95-Percent Confidence Interval		Test of Null Hypothesis (2-tail)	Retention Rate (percentage)		Number of Units Assigned[a]
		g	SE	Lower Limit	Upper Limit	Z-Value	Online	Face-to-Face	
Long and Jennings (2005) [Wave 2] [c]	The effect of technology and professional development on student achievement	+0.554	0.098	0.362	0.747	5.65***	Unknown	Unknown	6 teachers
Maki and Maki (2002)	Multimedia comprehension skill predicts differential outcomes of Web-based and lecture courses	+0.171	0.160	-0.144	0.485	1.06	91.01	88.10	155 students
Midmer, Kahan and Marlow (2006)	Effects of a distance learning program on physicians' opioid- and benzodiazepine-prescribing skills	+0.332	0.213	-0.085	0.750	1.56m	Unknown	Unknown	88 students
O'Dwyer, Carey and Kleiman (2007)	A study of the effectiveness of the Louisiana algebra I online course	+0.373	0.094	0.190	0.557	3.99***	88.51	64.4	Unknown[b]
Rockman et al. (2007) [Writing] [c]	ED PACE final report	-0.239	0.102	-0.438	-0.039	-2.34*	Unknown	Unknown	28 classrooms
Rockman et al. (2007) [Multiple-choice test] [c]	ED PACE final report	-0.146	0.102	-0.345	0.054	-1.43	Unknown	Unknown	28 classrooms
Schilling et al. (2006) [Search strategies] [c]	An interactive Web-based curriculum on evidence-based medicine: Design and effectiveness	+0.585	0.188	0.216	0.953	3.11**	68.66	59.62	Unknown
Schilling et al. (2006) [Quality of care calculation] [c]	An interactive Web-based curriculum on evidence-based medicine: Design and effectiveness	+0.926	0.183	0.567	1.285	5.05***	66.42	86.54	Unknown
Spires et al. (2001)	Exploring the academic self within an electronic mail environment	+0.571	0.357	-0.130	1.271	1.60	Unknown	100.00	31 students
Suter and Perry (1997)	Evaluation by electronic mail	+0.140	0.167	-0.188	0.468	0.84	Unknown	Unknown	Unknown

Exhibit 4b: Blended versus Face-to-Face (Category 2) Studies Included in the Meta-analysis (continued)

Authors	Title	Effect Size		95-Percent Confidence Interval		Test of Null Hypothesis (2-tail)	Retention Rate (percentage)		Number of Units Assigned[a]
		g	SE	Lower Limit	Upper Limit	Z-Value	Online	Face-to-Face	
Urban (2006)	A comparison of computer-based distance education and traditional tutorial sessions in supplemental instruction for students at-risk for academic difficulties	+0.264	0.192	-0.112	0.639	1.37	96.86	73.85	110 students
Zacharia (2007)	Comparing and combining real and virtual experimentation: An effort to enhance students' conceptual understanding of electric circuits	+0.570	0.216	0.147	0.993	2.64**	100	95.56	88 students

Exhibit reads: The effect size for the Aberson et al. (2003) study of an interactive tutorial on hypothesis testing was +0.58, which was not significantly different from 0.

*p < .05, ** p < .01, *** p < .001, *SE* = Standard error.

[a] This number represents the assigned units at study conclusion. It excludes units that attrited.

[b] The study involved 18 online classrooms from six districts and two private schools; the same six districts were asked to identify comparable face-to-face classrooms, but the study does not report how many of those classrooms participated.

[c] Two independent contrasts were contained in this article, which therefore appears twice in the table.

Test for Homogeneity

Both the Category 1 and Category 2 studies contrasted a condition with online elements with a condition of face-to-face instruction only. Analysts used the larger corpus of 50 effects that were either Category 1 or Category 2 to explore the influence of possible moderator variables.

The individual effect size estimates included in this meta-analysis ranged from a low of –0.80 (tendency for higher performance in the face-to-face condition) to a high of +1.11 (favoring online instruction). A test for homogeneity of effect size found significant differences across studies ($Q = 168.86$, $p < .0001$). Because of these significant differences in effect sizes, analysts investigated the variables that may have influenced the differing effect sizes.

Analyses of Moderator Variables

As noted in chapter 1, this meta-analysis has distinguished between *practice* variables, which can be considered part of intervention implementation, and *conditions*, which are status variables that are fairly impervious to outside influence. Relying on prior research, the research team identified variables of both types that might be expected to correlate with the effectiveness of online learning. The researchers also considered the potential influence of *study method* variables, which often vary with effect size; typically, more poorly controlled studies show larger effects. Each study in the meta-analysis was coded for these three types of variables—practice, status, and study method—using the coding categories shown in the appendix.

Many of the studies did not provide information about features considered to be potential moderator variables, a predicament noted in previous meta-analyses (see Bernard et al. 2004). Many of the reviewed studies, for example, did not indicate (a) whether or not the online instructor had received training in the method of instruction, (b) rates of attrition from the contrasting conditions and (c) contamination between conditions.

For some of the variables, the number of studies providing sufficient information to support categorization as to whether or not the feature was present was too small to support a meaningful analysis. Analysts identified those variables for which at least two contrasting subsets of studies, with each subset containing six or more study effects, could be constructed. In some cases, this criterion could be met by combining related feature codes; in a few cases, the inference was made that failure to mention a particular practice or technology (e.g., one-way video) denoted its absence. Practice, conditions and method variables for which study subsets met the size criterion were included in the search for moderator variables.

Practice Variables

Exhibit 5 shows the variation in effectiveness associated with 12 practice variables. Analysis of these variables addresses the third research question:

What practices are associated with more effective online learning?

Exhibit 5 and the two data exhibits that follow show significance results both for the various subsets of studies considered individually and for the test of the dimension used to subdivide the study sample (i.e., the potential moderator variable). For example, in the case of Computer-Mediated Communication With Peers, both the 17 contrasts in which students in the online condition had only asynchronous communication with peers and the 6 contrasts in which online students had both synchronous and asynchronous communication with peers are shown in the table. The two subsets had mean effect sizes of +0.27 and +0.17, respectively, and only the former was statistically different from 0. The *Q*-statistic of homogeneity tests whether the variability in effect sizes for these contrasts is associated with the type of peer communication available. The *Q*-statistic for Computer-Mediated Communication With Peers (0.32) is not statistically different from 0, indicating that the addition of synchronous communication with peers is not a significant moderator of online learning effectiveness.

The test of the moderator variable most central to this study—whether a blended online condition including face-to-face elements is associated with greater advantages over classroom instruction than is pure online learning—was discussed above. As noted there, the effect size for blended approaches contrasted against face-to-face instruction is larger than that for purely online approaches contrasted against face-to-face instruction. The other two practice variables included in the chapter 1 conceptual framework—learning experience type and synchronous versus asynchronous communication with the instructor—were tested in a similar fashion. The former was found to moderate significantly the size of the online learning effect ($Q = 6.19$, $p < .05$).[16] The mean effect size for collaborative instruction (+0.25) as well as for instructor-directed instruction (+0.39) were significantly positive whereas the mean effect size for independent, active online learning (+0.05) was not.[17]

Among the other 10 practices, which were not part of the conceptual model, none attained statistical significance.[18] The amount of time that students in the treatment condition spent on task compared with students in the face-to-face condition did approach statistical significance as a moderator of effectiveness ($Q = 3.62$, $p = .06$).[19] The mean effect size for studies with more time spent on task by online learners than learners in the control condition was +0.45 compared

[16] This contrast is not statistically significant ($p = .13$) when the five K-12 studies are removed from the analysis.

[17] Online experiences in which students explored digital artifacts and controlled the specific material they wanted to view were categorized as "independent" learning experiences.

[18] When the five K-12 studies are removed from the analysis, two additional practices are found to be statistically significant moderators of the effects of online learning – time spent on task and opportunities for face-to-face interactions with peers.

[19] Time on task as a moderator becomes statistically significant ($Q = 4.44$, $p < .05$) when the five K-12 studies are removed from the analysis.

Exhibit 5. Tests of Practices as Moderator Variables

Variable	Contrast	Number Studies	Weighted Effect Size	Standard Error	Lower Limit	Upper Limit	Q-Statistic
Pedagogy/ learning experience[a]	Instructor-directed (expository)	8	0.386**	0.120	0.150	0.622	6.19*
	Independent (active)	17	0.050	0.082	-0.110	0.210	
	Collaborative (interactive)	22	0.249***	0.075	0.102	0.397	
Computer-mediated communication with instructor[a]	Asynchronous only	16	0.239*	0.108	0.027	0.451	1.20
	Synchronous + Asynchronous	8	0.036	0.151	-0.259	0.331	
Computer-mediated communication with peers[a]	Asynchronous only	17	0.272**	0.091	0.093	0.450	0.32
	Synchronous + Asynchronous	6	0.168	0.158	-0.141	0.478	
Treatment duration[a]	Less than 1 month	19	0.140	0.089	-0.034	0.314	0.69
	More than 1 month	29	0.234***	0.069	0.098	0.370	
Media features[a]	Text-based only	14	0.208	0.111	-0.009	0.425	0.00
	Text + other media	32	0.200**	0.066	0.071	0.329	
Time on task[a]	Online > Face to Face	9	0.451***	0.113	0.229	0.673	3.62
	Same or Face to Face > Online	18	0.183*	0.083	0.020	0.346	
One-way video or audio	Present	14	0.092	0.091	-0.087	0.271	2.15
	Absent/Not reported	36	0.254***	0.062	0.133	0.375	
Computer-based instruction elements	Present	29	0.182**	0.065	0.054	0.311	0.25
	Absent/Not reported	21	0.234**	0.081	0.075	0.393	
Opportunity for face-to-face time with instructor	During instruction	21	0.298***	0.074	0.154	0.442	3.70
	Before or after instruction	11	0.050	0.118	-0.181	0.281	
	Absent/Not reported	18	0.150	0.091	-0.028	0.327	
Opportunity for face-to-face time with peers	During instruction	21	0.300***	0.072	0.159	0.442	5.20
	Before or after instruction	12	0.001	0.111	-0.216	0.218	
	Absent/Not reported	17	0.184*	0.093	0.001	0.367	
Opportunity to practice	Present	41	0.212***	0.056	0.102	0.322	0.15
	Absent/Not reported	9	0.159	0.124	-0.084	0.402	
Feedback provided	Present	23	0.204**	0.078	0.051	0.356	0.00
	Absent/Not reported	27	0.203**	0.070	0.066	0.339	

Exhibit reads: Studies in which time spent in online learning exceeded time in the face-to-face condition had a mean effect size of +0.45 compared with +0.18 for studies in which face-to-face learners had as much or more instructional time.

*$p < .05$, **$p < .01$, ***$p < .001$.

[a] The moderator analysis for this variable excluded studies that did not report information for this feature.

with +0.18 for studies in which the learners in the face-to-face condition spent as much or more time on task .

Condition Variables

The strategy to investigate whether study effect sizes varied with publication year, which was taken as a proxy for the sophistication of available technology, involved splitting the study sample into two nearly equal subsets by contrasting studies published between 1997 and 2003 against those published in 2004 through July 2008.

The studies were divided into three subsets of learner type: K–12 students, undergraduate students (the largest single group), and other types of learners (graduate students or individuals receiving job-related training). As noted above, the studies covered a wide range of subjects, but medicine and health care were the most common. Accordingly, these studies were contrasted against studies in other fields. Tests of these conditions as potential moderator variables addressed the study's fourth research question:

What conditions influence the effectiveness of online learning?

None of the three conditions tested emerged as a statistically significant moderator variable. In other words, for the range of student types for which studies are available, the effectiveness of online learning was equivalent in older and newer studies, with undergraduate and older learners, and in both medical and other subject areas. Exhibit 6 provides the results of the analysis of conditions.

Exhibit 6. Tests of Conditions as Moderator Variables

Variable	Contrast	Number of Contrasts	Weighted Effect Size	Standard Error	Lower Limit	Upper Limit	*Q*-Statistic
Year Published	1997–2003	13	0.195	0.105	-0.010	0.400	0.00
	2004 or after	37	0.203***	0.058	0.088	0.317	
Learner Type	K–12 students	7	0.1664	0.118	-0.065	0.397	3.25
	Undergraduate	21	0.309***	0.083	0.147	0.471	
	Graduate student/Other	21	0.100	0.084	-0.064	0.264	
Subject Matter	Medical/ Health care	16	0.205*	0.090	0.028	0.382	0.00
	Other	34	0.199**	0.062	0.0770	0.320	

Exhibit reads: The positive effect associated with online learning over face-to-face instruction was significant both for studies published between 1997 and 2003 and for those published in 2004 or later; the effect size does not vary significantly with period of publication.
*$p < .05$, **$p < .01$, ***$p < .001$.

Because of the *Evaluation of Evidence-Based Practices in Online Learning* study's emphasis on K-12 education, the online learning studies involving K-12 students were of particular interest. The meta-analysis includes seven contrasts from five studies of K-12 school students' online learning. Exhibit 7 describes these studies.

Given the small number of studies that addressed K-12 learners in the meta-analysis, attempts to test for statistical differences between the mean effect for K-12 learners and those for other types of learners should be viewed as merely suggestive. At +0.17, the average effect size for the seven contrasts involving K-12 learners appears similar to that for graduate and other students (+0.10) but less positive than that for undergraduates (+0.31). When learner type was tested as a moderator variable, however, the resulting Q-statistic was not significant.

Methods Variables

The advantage of meta-analysis is its ability to uncover general effects by looking across a range of studies that have operationalized the construct under study in different ways, studied it in different contexts, and used different methods and outcome measures. However, the inclusion of poorly designed and small-sample studies in the meta-analysis corpus poses concerns because doing so may give undue weight to spurious effects. Study methods variables were examined as potential moderators to explore this issue. The results are shown in Exhibit 8.

The influence of study sample size was examined by dividing studies into three subsets, according to the number of learners for which outcome data were collected. Sample size was not found to be a statistically significant moderator of online learning effects. Thus, there is no evidence that the inclusion of small-sample studies in the meta-analysis was responsible for the overall finding of a positive outcome for online learning.

Comparisons of the three designs deemed acceptable for this meta-analysis (random-assignment experiments, quasi-experiments with statistical control and crossover designs) indicate that study design is not significant as a moderator variable (see Exhibit 8). Moreover, in contrast with early meta-analyses in computer-based instruction, where effect size was inversely related to study design quality (Pearson et al. 2005), those experiments that used random assignment in the present corpus produced significant positive effects (+0.25, $p < .001$) while the quasi-experiments and crossover designs did not.

Exhibit 7. Studies of Online Learning Involving K–12 Students

The meta-analysis study corpus for this meta-analysis included five articles reporting on studies involving K–12 students. All of these studies compared student learning in a blended condition with student learning in a face-to-face condition. One of the studies (Long and Jennings 2005, Wave 1 study) was a randomized control trial and the others were quasi-experiments. One of the quasi-experiments (Rockman et al. 2007) provided two effect sizes that favored the face-to-face condition; the other studies provided five effects favoring the blended condition (with a range from +0.03 to +0.74).

Rockman et al. (2007) used a quasi-experimental matched comparison design to evaluate the effectiveness of Spanish courses offered to middle schools (seventh and eighth grades) through the West Virginia Virtual School. This virtual school program used a blended model of instruction that combined face-to-face and virtual instruction as well as paper and pencil and Web-based activities. The program was delivered by a three-member teacher team that included a lead teacher (a certified Spanish teacher) who was responsible for the design and delivery of the daily lesson plan and weekly phone conversations with each class; an adjunct teacher (a certified Spanish teacher) who provided content-related feedback by means of e-mail and voice-mail and who graded student tests and products; and a classroom facilitator (a certified teacher, but not a Spanish teacher) who guided students on site to ensure that they stayed on task and completed assignments on time. The hybrid Spanish course was offered to students in 21 schools that did not have the resources to provide face-to-face Spanish instruction. The students in the face-to-face group came from seven schools that matched the virtual schools with respect to average language arts achievement and school size. The study involved a total of 463 students.

Information needed to compute effect sizes was reported for two of the student learning measures used in the study. For the first of these, a multiple-choice test including subtests on oral and written comprehension of Spanish, the mean estimated effect was –0.15, and the difference between the two conditions was not statistically significant. The other measure was a test of students' writing ability, and the effect size for this skill was –0.24, with students receiving face-to-face instruction doing significantly better than those receiving the online blended version of the course.

Contrasting results were obtained in the other large-scale K–12 study, conducted by O'Dwyer, Carey and Kleiman (2007). These investigators used a quasi-experimental design to compare the learning of students participating in the Louisiana Algebra I Online initiative with the learning of students in comparison classrooms that were "similar with regard to mathematics ability, environment, and size, but where teachers used traditional 'business as usual' approaches to teaching algebra" (p. 293). Like the West Virginia Virtual School program, this initiative used a blended model of instruction that combined face-to-face and Web-based activities with two teachers: one in class and the other online. Matched pre- and posttest scores on researcher-developed multiple-choice tests were collected from a total of 463 students (231 from the treatment group, 232 from the comparison group) from multiple schools and school districts. An effect size of +0.37 was obtained, with online students performing better than their peers in conventional classrooms.

Exhibit 7. Studies of Online Learning Involving K-12 Students (continued)

Long and Jennings (2005) examined whether the performance of eighth-grade students whose teachers integrated the use of the *Pathways to Freedom* Electronic Field Trips—an online collection of interactive activities designed by Maryland Public Television—improved compared with performance of students whose teachers taught the same content without the online materials. The study provided two sets of analyses from two waves of data collection, yielding two independent effect sizes. The first set of analyses involved the data from nine schools in two Maryland districts. Schools were assigned randomly to conditions. Teachers in both conditions covered the same learning objectives related to slavery and the Underground Railroad, with the treatment teachers using the *Pathways to Freedom* Electronic Field Trips materials. A small effect size of +0.03 favoring the online condition was computed from change scores on researcher-developed multiple-choice tests administered to 971 students.

Long and Jennings' (2005, wave 2) second study involved a subset of teachers from one of the two participating districts, which was on a semester schedule. The teachers from this district covered the same curriculum twice during the year for two different sets of students. The gain scores of 846 students of six teachers (three treatment teachers and three control teachers) from both semesters were collected. Regression analysis indicated an effect size of +0.55 favoring the online conditions. This study also looked into the maturation effects of teachers' using the online materials for the second time. As hypothesized, the results showed that the online materials were used more effectively in the second semester.

Sun, Lin and Yu (2008) conducted a quasi-experimental study to examine the effectiveness of a virtual Web-based science lab with 113 fifth-grade students in Taiwan. Although both treatment and control groups received an equal number of class hours and although both groups conducted manual experiments, students in the treatment condition used the virtual Web-based science lab for part of their lab time. The Web-based lab enabled students to conduct virtual experiments while teachers observed student work and corrected errors online. The control group students conducted equivalent experiments using conventional lab equipment. Matched pre- and posttest scores on researcher-developed assessments were collected for a total of 113 students (56 from the treatment group and 57 from the comparison group) in four classrooms from two randomly sampled schools. An effect size of +0.26 favoring the virtual lab condition was obtained from analysis of covariance results, controlling for pretest scores.

A small-scale quasi-experiment was conducted by Englert et al. (2007). This study examined the effectiveness of a Web-based writing support program with 35 elementary-age students from six special education classrooms across five urban schools. Students in the treatment group used a Web-based program that supported writing performance by prompting attention to the topical organization and structure of ideas during the planning and composing phases of writing. Control students used similar writing tools provided in traditional paper-and-pencil formats. Pre- and posttests of student writing, scored on a researcher-developed rubric, were used as outcome measures. An effect size of +0.74 favoring the online condition was obtained from an analysis of covariance controlling for writing pretest scores.

Exhibit 8. Tests of Study Features as Moderator Variables

Variable	Contrast	Number of Studies	Weighted Effect Size	Standard Error	Lower Limit	Upper Limit	*Q*-Statistic
Sample size	Fewer than 35	11	0.203	0.139	-0.069	0.476	0.01
	From 35 to 100	20	0.209*	0.086	0.039	0.378	
	More than 100	19	0.199**	0.072	0.058	0.339	
Type of knowledge tested[a]	Declarative	12	0.180	0.097	-0.010	0.370	0.37
	Procedural/ Procedural and declarative	30	0.239***	0.068	0.106	0.373	
	Strategic knowledge	5	0.281	0.168	-0.047	0.610	
Study design	Random assignment control	32	0.249***	0.065	0.122	0.376	1.50
	Quasi-experimental design with statistical control	13	0.108	0.095	-0.079	0.295	
	Crossover design	5	0.189	0.158	-0.120	0.499	
Unit of assignment to conditions[a]	Individual	32	0.169*	0.066	0.040	0. 298	4.73.
	Class section	7	0.475***	0.139	0.202	0.748	
	Course/School	9	0.120	0.103	-0.083	0.323	
Instructor equivalence[a]	Same instructor	20	0.176*	0.078	0.024	0.329	0.73
	Different instructor	19	0.083	0.077	-0.067	0.233	
Equivalence of curriculum/ instruction[a]	Identical/ Almost identical	29	0.130*	0.063	0.007	0.252	6.85**
	Different/ Somewhat different	17	0.402***	0.083	0.239	0.565	

Exhibit reads: The average effect size was significantly positive for studies with a sample size of less than 35 as well as for those with 35 to 100 and those with a sample size larger than 100; the weighted average effect did not vary with size of the study sample.

*$p < .05$, **$p < .01$, ***$p < .001$.

[a] The moderator analysis excluded some studies because they did not report information about this feature.

Effect sizes do not vary depending on whether or not the same instructor or instructors taught in the face-to-face and online conditions ($Q = 0.73$, $p > .05$). The average effect size for the 20 contrasts in which instructors were the same across conditions was +0.18, $p < .05$. The average effect size for contrasts in which instructors varied across conditions was +0.08, $p > .05$. The only study method variable that proved to be a significant moderator of effect size was comparability of the instructional materials and approach for treatment and control students.

The analysts coding study features examined the descriptions of the instructional materials and the instructional approach for each study and coded them as "identical," "almost identical," "different" or "somewhat different" across conditions. Adjacent coding categories were combined (creating the two study subsets Identical/Almost Identical and Different/Somewhat Different) to test Equivalence of Curriculum/Instruction as a moderator variable. Equivalence of Curriculum/Instruction was a significant moderator variable ($Q = 6.85$, $p < .01$). An examination of the study subgroups shows that the average effect for studies in which online learning and face-to-face instruction were described as identical or nearly so was +0.13, $p < .05$, compared with an average effect of +0.40 ($p < .001$) for studies in which curriculum materials and instructional approach varied across conditions.

The moderator variable analysis for aspects of study method also found additional patterns that did not attain statistical significance but that should be re-tested once the set of available rigorous studies of online learning has expanded. The type of learning outcome tested, for example, may influence the magnitude of effect sizes. Twelve studies measured declarative knowledge outcomes only, typically through multiple-choice tests. A larger group of studies (30) looked at students' ability to perform a procedure, or they combined procedural and declarative knowledge outcomes in their learning measure. Five studies used an outcome measure that focused on strategic knowledge. (Three studies did not describe their outcome measures in enough detail to support categorization.) Among the subsets of studies, the average effect for studies that included procedural knowledge in their learning outcome measure (effect size of +0.24) and that for studies that measured strategic knowledge (effect size of +0.28) appeared larger than the mean effect size for studies that used a measure of declarative knowledge only (+0.18). Even so, the Type of Knowledge Tested was not a significant moderator variable ($Q = 0.37$, $p > .05$).

4. Narrative Synthesis of Studies Comparing Variants of Online Learning

This chapter presents a narrative summary of Category 3 studies—those that examined the learning effects of variations in online practices such as different versions of blended instruction or online learning with and without immediate feedback to the learner. The literature search and screening (described in chapter 2) identified 84 Category 3 studies reported in 79 articles.[20]

Within the set of Category 3 studies, five used K–12 students as subjects and 10 involved K–12 teacher education or professional development. College undergraduates constituted the most common learner type (see Exhibit 9). All Category 3 studies involved formal education. Course content for Category 3 studies covered a broad range of subjects, including observation skills, understanding Internet search engines, HIV/AIDS knowledge and statistics.

When possible, the treatment manipulations in Category 3 studies were coded using the practice variable categories that were used in the meta-analysis to facilitate comparisons of findings between the meta-analysis and the narrative synthesis. No attempt was made to statistically combine Category 3 study results, however, because of the wide range of conditions compared in the different studies.

Exhibit 9. Learner Types for Category 3 Studies

Educational Level	Number of Studies
K–12	5
Undergraduate	37
Graduate	4
Medical[a]	18
Teacher professional development[b]	10
Adult training	4
Other[c]	4
Not available	2
Total	**84**

Exhibit reads: K–12 students were the learners in 5 of the 84 studies of alternative online practices.
[a] The medical category spans undergraduate and graduate educational levels and includes nursing and related training.
[b] Teacher professional development includes preservice and inservice training.
[c] The Other category includes populations consisting of a combination of learner types such as student and adult learners or undergraduate and graduate learners.

[20] Some articles contained not only contrasts that fit the criteria for Category 1 or 2 but also contrasts that fit Category 3. The appropriate contrasts between online and face-to-face conditions were used in the meta-analysis; the other contrasts were reviewed as part of the Category 3 narrative synthesis presented here.

Blended Compared With Pure Online Learning

The meta-analysis of Category 1 and 2 studies described in chapter 3 found that effect sizes were larger for studies that compared blended learning conditions with face-to-face instruction than for studies that compared purely online learning with face-to-face instruction. Another way to investigate the same issue is by conducting studies that incorporate both blended and purely online conditions to permit direct comparisons of their effectiveness.

The majority of the 10 Category 3 studies that directly compared purely online and blended learning conditions found no significant differences in student learning. Seven studies found no significant difference between the two, two found statistically significant advantages for purely online instruction, and one found an advantage for blended instruction. The descriptions of some of these studies, provided below, make it clear that although conditions were labeled as "blended" or "purely online" on the basis of their inclusion or exclusion of face-to-face interactions, conditions differed in terms of content and quality of instruction. Across studies, these differences in the nature of purely online and blended conditions very likely contributed to the variation in outcomes.

Keefe (2003), for example, contrasted a section of an organizational behavior course that received lectures face-to-face with another section that watched narrated PowerPoint slides shown online or by means of a CD-ROM. Both groups had access to e-mail, online chat rooms, and threaded discussion forums. All course materials were delivered electronically to all students at the same time. On the course examination, students in the purely online section scored almost 8 percent lower than those receiving face-to-face lectures in addition to the online learning activities. Keefe's was the only study in the review that found a significant decrement in performance for the condition without face-to-face instructional elements.

Poirier and Feldman (2004) compared a course that was predominantly face-to-face but also used an online discussion board with a course taught entirely online. Students in the predominantly face-to-face version of the course were required to participate in three online discussions during the course and to post at least two comments per discussion to an online site; the site included content, communication and assessment tools. In the purely online version of the course, students and the instructor participated in two online discussions each week. Poirier and Feldman found a significant main effect favoring the purely online course format for examination grades but no effect on student performance on writing assignments.

Campbell et al. (2008) compared a blended course (in which students accessed instruction online but attended face-to-face discussions) with a purely online course (in which students accessed instruction and participated in discussions online). Tutors were present in both discussion formats. Students were able to select the type of instruction they wanted, blended or online. Mean scores for online discussion students were significantly higher than those for the face-to-face discussion group.

As a group, these three studies suggest that the relative efficacy of blended and purely online learning approaches depends on the instructional elements of the two conditions. For the most part, these studies did not control instructional content within the two delivery conditions (blend

of online and face-to-face versus online only). For example, the lecturer in the Keefe (2003) study may have covered material not available to the students reviewing the lecture's PowerPoint slides online. Alternately, in the Poirier and Feldman (2004) study, students interacting with the instructor in two online discussions a week may have received more content than did those receiving face-to-face lectures.

Davis et al. (1999) attempted to equate the content delivered in their three class sections (online, traditional face-to-face, and a blended condition in which students and instructor met face-to-face but used the online modules). Students in an educational technology course were randomly assigned to one of the three sections. No significant differences among the three conditions were found in posttest scores on a multiple-choice test.

An additional six studies contrasting purely online conditions and blended conditions (without necessarily equating learning content across conditions) also failed to find significant differences in student learning. Ruchti and Odell (2002) compared test scores from two groups of students taking a course on elementary science teaching methods. One group took online modules; the other group received instruction in a regular class, supplemented with an online discussion board and journal (also used in the online course condition). No significant difference between the groups was found.

Beile and Boote (2002) compared three groups: one with face-to-face instruction alone, another with face-to-face instruction and a Web-based tutorial, and a third with Web-based instruction and the same Web-based tutorial. The final quiz on library skills indicated no significant differences among conditions.

Gaddis et al. (2000) compared composition students' audience awareness between a blended course and a course taught entirely online. The same instructor taught both groups, which also had the same writing assignments. Both groups used networked computers in instruction, in writing and for communication. However, the "on campus" group met face-to-face, giving students the opportunity to communicate in person, whereas the "off campus" group met only online. The study found no significant difference in learner outcomes between the two groups.

Similarly, Caldwell (2006) found no significant differences in performance on a multiple-choice test between undergraduate computer science majors enrolled in a blended course and those enrolled in an online course. Both groups used a Web-based platform for instruction, which was supplemented by a face-to-face lab component for the blended group.

Scoville and Buskirk (2007) examined whether the use of traditional or virtual microscopy would affect learning outcomes in a medical histology course. Students were assigned to one of four sections: (a) a control section where learning and testing took place face-to-face, (b) a blended condition where learning took place virtually and the practical examination took place face-to-face, (c) a second blended condition where learning took place face-to-face and testing took place virtually, and (d) a fully online condition. Scoville and Buskirk found no significant differences in unit test scores by learning groups.

Finally, McNamara et al. (2008) studied the effectiveness of different approaches to teaching a weight-training course. They divided students into three groups: a control group that received

face-to-face instruction, a blended group that received a blend of online and face-to-face instruction, and a fully online group. The authors did not find a significant main effect for group type.[21]

Thus, as a group, these studies do not provide a basis for choosing online versus blended instructional conditions

Media Elements

Eight studies in the Category 3 corpus compared online environments using different media elements such as one-way video (Maag 2004; McKethan et al. 2003; Schmeeckle 2003; Schnitman 2007; Schroeder 2006; Schutt 2007; Tantrarungroj 2008; Zhang et al. 2006). *Seven of the eight studies found no significant differences among media combinations.* In the study that found a positive effect from enhanced media features, Tantrarungroj (2008) compared two instructional approaches for teaching a neuroscience lesson to undergraduate students enrolled in computer science classes. The author contrasted an experimental condition in which students were exposed to online text with static graphics and embedded video with a control condition in which students did not have access to the streaming video. Tantrarungroj found no significant difference in grades for students in the two conditions on a posttest administered immediately after the course; however, the treatment group scored significantly higher on a knowledge retention test that was administered 4 weeks after the intervention.

The other seven studies found no effect on learning from adding additional media to online instruction. For example, Schnitman (2007) sought to determine whether enhancing text with graphics, navigation options, and color would affect learning outcomes. The author randomly assigned students to one of two conditions in a Web-based learning interface; the control group accessed a plain, text-based interface, and the treatment group accessed an enhanced interface that featured additional graphics, navigational options, and an enhanced color scheme. Schnitman found no significant differences in learning outcomes between the treatment and control groups.

The fact that the majority of studies found no significant difference across media types is consistent with the theoretical position that the medium is simply a carrier of content and is unlikely to affect learning per se (Clark 1983, 1994). A study by Zhang et al. (2006) suggests that the way in which a medium is used is more important than merely having access to it. Zhang et al. found that the effect of video on learning hinged on the learner's ability to control the video ("interactive video"). The authors used four conditions: traditional face-to-face and three online environments—interactive video, noninteractive video, and nonvideo. Students were randomly assigned to one of the four groups. Students in the interactive video group performed significantly better than the other three groups. There was no statistical difference between the online group that had noninteractive video and the online group that had no video.

[21] However, in tests of cognitive knowledge and strength, both the control and blended sections showed significant improvements, whereas the fully online section showed no significant pre- to posttest growth for either outcome.

In summary, many researchers have hypothesized that the addition of images, graphics, audio, video or some combination would enhance student learning and positively affect achievement. However, the majority of studies to date have found that these media features do not affect learning outcomes significantly.

Learning Experience Type

Other Category 3 studies manipulated different features of the online learning environment to investigate the effects of learner control or type of learning experience. *The learning experience studies provide some evidence that suggests an advantage for giving learners an element of control over the online resources with which they engage; however, the studies' findings are mixed with respect to the relative effectiveness of the three learning experience types in the conceptual framework presented in chapter 2.*

Four studies (Cavus et al. 2007; Dinov, Sanchez and Christou 2008; Gao and Lehman 2003; Zhang 2005) provide preliminary evidence supporting the hypothesis that conditions in which learners have more control of their learning (either active or interactive learning experiences in our conceptual framework) produce larger learning gains than do instructor-directed conditions (expository learning experiences). Three other studies failed to find such an effect (Cook et al. 2007; Evans 2007; Smith 2006).

Zhang (2005) reports on two studies comparing expository learning with active learning, both of which found statistically positive results in favor of active learning. Zhang manipulated the functionality of a Web course to create two conditions. For the control group, video and other instruction received over the Web had to be viewed in a specified order, videos had to be viewed in their entirety (e.g., a student could not fast forward) and rewinding was not allowed. The treatment group could randomly access materials, watching videos in any sequence, rewinding them and fast forwarding through their content. Zhang found a statistically significant positive effect in favor of learner control over Web functionality (see also the Zhang et al. 2006 study described above). Gao and Lehman (2003) found that students who were required to complete a "generative activity" in addition to viewing a static Web page performed better on a test about copyright law than did students who viewed only the static Web page. Cavus, Uzonboylu and Ibrahim (2007) compared the success rates of students learning the Java programming language who used a standard collaborative tool with the success rate of those who used an advanced collaborative tool that allowed compiling, saving and running programs inside the tool. The course grades for students using the advanced collaborative tool were higher than those of students using the more standard tool. Similarly, Dinov, Sanchez and Christou (2008) integrated tools from the Statistics Online Computational Resource in three courses in probability and statistics. For each course, two groups were compared: one group of students received a "low-intensity" experience that provided them with access to a few online statistical tools; the other students received a "high-intensity" condition with access to many online tools for acting on data. Across the three classes, pooling all sections, students in the more active, high-intensity online tool condition demonstrated better understanding of the material on mid-term and final examinations than did the other students.

These studies that found positive effects for learner control and nondidactic forms of instruction are counterbalanced by studies that found mixed or null effects from efforts to provide a more

active online learning experience. Using randomly assigned groups of nurses who learned about pain management online, Smith (2006) altered the instructional design to compare a text-based, expository linear design with an instructional design involving participant problem solving and inquiry. No significant difference was found between the two groups in terms of learning outcomes. Cook et al. (2007) found no differences in student learning between a condition with end-of-module review questions that required active responses and a condition with expository end-of-module activities. Evans (2007) explored the effects of more and less expository online instruction for students learning chemistry lab procedures. After asking students to complete an online unit that was either text-based or dynamic and interactive, Evans found that SAT score and gender were stronger predictors of student performance on a posttest with conceptual and procedural items than was the type of online unit to which students were exposed.

Golanics and Nussbaum (2008) examined the effect of "elaborated questions" and "maximizing reasons" prompts on students' ability to construct and critique arguments. Students were randomly divided into groups of three; each group engaged in asynchronous discussions. Half of the groups received "elaborated questions," which explicitly instructed them to think of arguments and counterarguments, whereas the other half of the groups viewed unelaborated questions. In addition, half of the groups randomly received prompts to provide justifications and evidence for their arguments (called the "maximizing reasons" condition); half of the groups did not receive those prompts. Elaborated questions stimulated better-developed arguments, but maximizing reasons instructions did not.

Chen (2007) randomly assigned students in a health-care ethics class to one of three Web-based conditions: (a) a control group that received online instruction without access to an advanced organizer; (b) a treatment group that studied a text-based advanced organizer before online instruction; and (c) a second treatment group that reviewed an advanced, Flash-based concept map organizer before engaging in online learning.[22] The authors hypothesized that both the advanced organizer and the concept map would help students access relevant prior knowledge and increase their active engagement with the new content. Contrary to expectations, Chen found no significant differences in learning achievement across the three groups.

Suh (2006) examined the effect of guiding questions on students' ability to produce a good educational Web site as required in an online educational technology course. Students in the guiding-question condition received questions through an electronic discussion board and were required to read the questions before posting their responses. E-mails and online postings reminded them to think about the guiding questions as they worked through the problem scenario. Guiding questions were found to enhance the performance of students working alone, but they did not produce benefits for students working in groups. One possible explanation offered by the author is that students working in groups may scaffold each other's work, hence reducing the benefit derived from externally provided questions.

[22] Flash animations are created using Flash software from Adobe; a concept map is a graphic depiction of a set of ideas and the linkages among them.

Computer-Based Instruction

The advantage of incorporating elements that are generally found in stand-alone computer-based instruction into online learning seems to depend on the nature of the contrasting conditions. Quizzes, simulations and individualized instruction, all common to stand-alone computer-based instruction, appear to vary in their effectiveness when added to an online learning environment.

Online Quizzes

Research on incorporating quizzes into online learning does not provide evidence that the practice is effective. The four studies that examined the effectiveness of online quizzes (Lewis 2002; Maag 2004; Stanley 2006; Tselios et al. 2001) had mixed findings. Maag (2004) and Stanley (2006) found no advantage for the inclusion of online quizzes. Maag included online quizzes in a treatment condition that also provided students with online images, text and some animation; the treatment group was compared with other groups, which differed both in the absence of online quizzes and in terms of the media used (one had the same text and images delivered online, one had printed text only, and one had printed text plus images). Maag found no significant difference between the online group that had the online quizzes and the online group that did not. Stanley (2006) found that outcomes for students taking weekly online quizzes did not differ statistically from those for students who completed homework instead.

Two other studies suggested that whether or not quizzes positively affect learning may depend on the presence of other variables. Lewis (2002) grouped students into two cohorts. For six modules, Group 1 took online quizzes and Group 2 participated in online discussions. For six other modules, the groups switched so that those who had been taking the online quizzes participated in online discussions and vice versa. When Group 1 students took the online quizzes, they did significantly better than those participating in discussions, but no difference was found between the groups when Group 2 took the online quizzes in the other six modules. The researchers interpreted this interaction between student group and condition in terms of the degree of interactivity in the online discussion groups. Group 1 was more active in the online discussions, and the authors suggested that this activity mitigated any loss in learning otherwise associated with not taking quizzes.

Tselios et al. (2001) suggest that the software platform used to deliver an online quiz may affect test performance. In their study, students completing an online quiz in WebCT performed significantly better than students taking the online quiz on a platform called IDLE. The educational content in the two platforms was identical and their functionality was similar; however, they varied in the details of their user interfaces.

Simulations

The results of three studies exploring the effects of including different types of online simulations were modestly positive. Two of the studies indicated a positive effect from including an online simulation; however, one study found no significant difference. In an online module on information technology for undergraduate psychology students, Castaneda (2008) contrasted two simulation conditions (one provided a simulation that students could explore as they chose, and the other guided the students' interaction with the simulation, providing some feedback and

expository material) with a condition that included no simulation. Castaneda also manipulated the sequencing of instructional activities, with the interaction with the simulation coming either before or after completion of the expository portion of the instructional module. Knowledge gains from pre- to posttest were greater for students with either type of simulation, provided they were exposed to it after, rather than before, the expository instruction.

Hibelink (2007) explored the effectiveness of using two-dimensional versus three-dimensional images of human anatomy in an online undergraduate human anatomy lab. The group of students that used three-dimensional images had a small, but significant advantage in identifying anatomical parts and spatial relationships. Contrasting results were obtained by Loar (2007) in an examination of the effects of computer-based case study simulations on students' diagnostic reasoning skills in nurse practitioner programs. All groups received identical online lectures, followed by an online text-based case study for one group and by completion of a computer-simulated case study for the other. No difference was found between the group receiving the case simulation versus that receiving the text-based version of the same case.

Individualized Instruction

The online learning literature has also explored the effects of using computer-based instruction elements to individualize instruction so that the online learning module or platform responds dynamically to the participant's questions, needs or performance. *There were only two online learning studies of the effects of individualizing instruction, but both found a positive effect.* Nguyen (2007) compared the experiences of people learning to complete tax preparation procedures, contrasting those who used more basic online training with those who used an enhanced interface that incorporated a context-sensitive set of features, including integrated tutorials, expert systems, and content delivered in visual, aural and textual forms. Nguyen found that this combination of enhancements had a positive effect.

Grant and Courtoreille (2007) studied the use of post-unit quizzes presented either as (a) fixed items that provided feedback only about whether or not the student's response was correct or (b) post-unit quizzes that gave the student the opportunity for additional practice on item types that had been answered incorrectly. The response-sensitive version of the tutorial was found to be more effective than the fixed-item version, resulting in greater changes between pre- and posttest scores.

Supports for Learner Reflection

Nine studies (Bixler 2008; Chang 2007; Chung, Chung and Severance 1999; Cook et al. 2005; Crippen and Earl 2007; Nelson 2007; Saito and Miwa 2007; Shen, Lee and Tsai 2007; Wang et al. 2006) examined the degree to which promoting aspects of learner reflection in a Web-based environment improved learning outcomes. *These studies found that a tool or feature prompting students to reflect on their learning was effective in improving outcomes.*

For example, Chung, Chung and Severance (1999) examined how computer prompts designed to encourage students to use self-explanation and self-monitoring strategies affected learning, as measured by students' ability to integrate ideas from a lecture into writing assignments. Chung et

al. found that students in the group receiving the computer prompts integrated and elaborated a significantly higher number of the concepts in their writing than did those in the control group.

In a quasi-experimental study of Taiwan middle school students taking a Web-based biology course, Wang et al. (2006) found that students in the condition using a formative online self-assessment strategy performed better than those in conditions using traditional tests, whether the traditional tests were online or administered in paper-and-pencil format. In the formative online assessment condition, when students answered an item incorrectly, they were told that their response was not correct, and they were given additional resources to explore to find the correct answer. (They were not given the right answer.) This finding is similar to that of Grant and Courtoreille (2007) described above.

Cook et al. (2005) investigated whether the inclusion of "self-assessment" questions at the end of modules improved student learning. The study used a randomized, controlled, crossover trial, in which each student took four modules, two with the self-assessment questions and two without. The order of modules was randomly assigned. Student performance was statistically higher on tests taken immediately after completion of modules that included self-assessment questions than after completion of those without such questions—an effect that the authors attributed to the stimulation of reflection. This effect, however, did not persist on an end-of-course test, on which all students performed similarly.

Shen, Lee and Tsai (2007) found a combination of effects for self-regulation and opportunities to learn through realistic problems. They compared the performance of students who did and did not receive instruction in self-regulation learning strategies such as managing study time, goal-setting and self-evaluation. The group that received instruction in self-regulated learning performed better in their online learning.

Bixler (2008) examined the effects of question prompts asking students to reflect on their problem-solving activities. Crippen and Earl (2007) investigated the effects of providing students with examples of chemistry problem solutions and prompts for students to provide explanations regarding their work. Chang (2007) added a self-monitoring form for students to record their study time and environment, note their learning process, predict their test scores and create a self-evaluation. Saito and Miwa (2007) investigated the effects of student reflection exercises during and after online learning activities. Nelson (2007) added a learning guidance system designed to support a student's hypothesis generation and testing processes without offering direct answers or making judgments about the student's actions. In all of these studies, the additional reflective elements improved students' online learning.

Overall, the available research evidence suggests that promoting self-reflection, self-regulation and self-monitoring leads to more positive online learning outcomes. Features such as prompts for reflection, self-explanation and self-monitoring strategies have shown promise for improving online learning outcomes.

Moderating Online Groups

Organizations providing or promoting online learning generally recommend the use of instructors or other adults as online moderators, but research support for the effects of this practice on student learning is mixed. A study by Bernard and Lundgren-Cayrol (2001) suggests that instructor moderation may not improve learning outcomes in all contexts. The study was conducted in a teacher education course on educational technology in which the primary pedagogical approach was collaborative, project-based learning. Students in the course were randomly assigned to groups receiving either low or high intervention on the part of a moderator and composed of either random or self-selected partners. The study did not find a main effect for moderator intervention. In fact, the mean examination scores of the low-moderation, random-selection groups were significantly higher than those of the other groups. A study by De Wever, Van Winckel and Valcke (2008) also found mixed effects resulting from instructor moderation. This study was conducted during a clinical rotation in pediatrics in which knowledge of patient management was developed through case-based asynchronous discussion groups. Researchers used a crossover design to create four conditions based on two variables: the type of moderator (instructor moderator versus student moderator) and the presence of a developer of alternatives for patient management (assigned developer versus no assigned developer). The presence of a course instructor as moderator was found not to improve learning outcomes significantly. When no assigned developer of alternatives was assigned, the two moderator conditions performed equivalently. When a developer of alternatives was specified, the student-moderated groups performed significantly better than the instructor-moderated groups.

Alternately, Zhang (2004) found that an externally moderated group scored significantly higher on problems calling for use of statistical knowledge and problem-solving skills than a peer-controlled group on both well- and ill-structured problems. Zhang's study compared the effectiveness of peer versus instructor moderation of online asynchronous collaboration. Students were randomly assigned to one of two groups. One group had a "private" online space where students entirely controlled discussion. The other group's discussion was moderated by the instructor, who also engaged with students through personal e-mails and other media.

Scripts for Online Interaction

Four Category 3 studies investigated alternatives to human moderation of online discussion in the form of "scaffolding" or "scripts" designed to produce more productive online interaction. *The majority of these studies indicated that the presence of scripts to guide interactions among groups learning together online did not appear to improve learning outcomes.*

The one study that found positive student outcomes for learners who had been provided scripts was conducted by Weinberger et al. (2005). These researchers created two types of scripts: "epistemic scripts," which specified how learners were to approach an assigned task and guided learners to particular concepts or aspects of an activity, and "social scripts," which structured how students should interact with each other through methods such as gathering information from each other by asking critical questions. They found that social scripts improved performance on tests of individual knowledge compared with a control group that participated in online discussions without either script (whether or not the epistemic script was provided).

The remaining three studies that examined the effect of providing scripts or scaffolds for online interaction found no significant effect on learning (Choi, Land and Turgeon 2005; Hron et al. 2000; Ryan 2007). Hron et al. (2000) used an experimental design to compare three groups: (a) a control that received no instructions regarding a 1-hour online discussion, (b) a group receiving organizing questions to help structure their online communication and (c) a group receiving both the organizing questions and rules for discussion. The discussion rules stated that group members should discuss only the organizing questions; that discussion of one question had to be completed before the next discussion was begun; that the discussion needed to be structured as an argument, with claims justified and alternative viewpoints considered; and that all participants should take turns moderating the discussion and making sure that the discussion adhered to the rules. Hron et al. found statistically significant differences across conditions in the content and coherence of student postings, but no difference across the three groups in terms of knowledge acquisition as measured by a multiple-choice test.

Ryan's study (2007) reached conclusions similar to those of Hron et al. Ryan hypothesized that exposure to collaborative tools would affect student performance. He compared two groups of middle school students: a treatment group, which engaged in online learning that included interaction with instructors and peers using online collaboration tools, and a control group, which did not have access to or instruction in the use of collaboration tools. Like Hron et al., Ryan found no significant difference in academic performance between the two groups of online students.

Choi, Land and Turgeon (2005) used a time-series control-group design to investigate the effects of providing online scaffolding for generating questions to peers during online group discussions. Although scaffolds were found to increase the number of questions asked, they did not affect question quality or learner outcomes.

In summary, mechanisms such as scaffolds or scripts for student group interaction online have been found to influence the way students engage with each other and with the online material, but have not been found to improve learning.

Delivery Platform

Several platform options are available for online learning—an exclusively Web-based environment or e-mail or mobile phone. The alternative platforms can be used as primary delivery channels or as supplements to Web-based instruction. Neither of the two studies that addressed this issue found significant differences across delivery platforms. Shih (2007) investigated whether student groups who accessed online materials by means of mobile phone demonstrated significantly different learning outcomes from groups who did so using a traditional computer; the author found no statistical difference between the two groups. Similarly, Kerfoot (2008) compared the effects of receiving course materials and information through a series of e-mails spaced out over time versus accessing the online materials all at once by means of a traditional Web-site and found no statistical difference.

Overall, the controlled studies are too few to support even tentative conclusions concerning the learning effects of using alternative or multiple delivery platforms for online learning.

Summary

This narrative review has illustrated the many variations in online, individual and group, and synchronous and asynchronous activities that can be combined in a course or instructional intervention. The number of Category 3 studies concerning any single practice was insufficient to warrant a quantitative meta-analysis, and the results varied to such an extent that only tentative, rather than firm, conclusions can be drawn about promising online learning practices.

The direct comparison of blended and purely online conditions in 10 studies produced mostly null results, tempering what appeared to be an advantage of blended compared with purely online instruction in the moderator variable analysis that was conducted as part of the meta-analysis presented in chapter 3. Although a fair number of Category 3 studies contrasted these two versions of online learning, few equated instructional content or activities across conditions, making it difficult to draw conclusions.

With respect to incorporation of multiple media, the evidence available in the Category 3 studies suggests that inclusion of more media in an online application does not enhance learning when content is controlled, but some evidence suggests that the learner's ability to control the learning media is important (Zhang 2005; Zhang et al. 2006). Alternately, the set of studies using various manipulations to try to stimulate more active engagement on the part of online learners (such as use of advanced organizers, conceptual maps, or guiding questions) had mostly null results.

The clearest recommendation for practice that can be made on the basis of the Category 3 synthesis is to incorporate mechanisms that promote student reflection on their level of understanding. A dozen studies have investigated what effects manipulations that trigger learner reflection and self-monitoring of understanding have on individual students' online learning outcomes. Ten of the studies found that the experimental manipulations offered advantages over online learning that did not provide the trigger for reflection.

Another set of studies explored features usually associated with computer-based instruction, including the incorporation of quizzes, simulations, and techniques for individualizing instruction. The providing of simple multiple-choice quizzes did not appear to enhance online learning. The incorporation of simulations produced positive effects in two out of three studies (Castaneda 2008; Hibelink 2007). Individualizing online learning by dynamically generating learning content based on the student's responses was found to be effective in the two studies investigating this topic (Grant and Courtoreille 2007; Nguyen 2007).

Attempts to guide the online interactions of groups of learners were less successful than the use of mechanisms to prompt reflection and self-assessment on the part of individual learners. Some researchers have suggested that students who learn in online groups provide scaffolds for one another (Suh 2006).

Finally, readers should be cautioned that the literature on alternative online learning practices has been conducted for the most part by professors and other instructors who are conducting research using their own courses. Moreover, the combinations of technology, content and activities used in different experimental conditions have often been ad hoc rather than theory based. As a result, the field lacks a coherent body of linked studies that systematically test theory-based approaches in different contexts.

5. Discussion and Implications

The meta-analysis reported here differs from prior meta-analyses of distance learning in several important respects:

- Only studies of Web-supported learning have been included.
- All effects have been based on objective measures of learning.
- Only studies with controlled designs that met minimum quality criteria have been included.

The corpus of 50 effect sizes extracted from 45 studies meeting these criteria was sufficient to demonstrate that in recent applications, online learning has been modestly more effective, on average, than the traditional face-to-face instruction with which it has been compared. It should be noted, however, that this overall effect can be attributed to the advantage of blended learning approaches over instruction conducted entirely face-to-face. Of the 11 individual studies with significant effects favoring the online condition, 9 used a blended learning approach.

The test for homogeneity of effects found significant variability in the effect sizes for the different online learning studies, justifying a search for moderator variables that could explain the differences in outcomes. The moderator variable analysis found only three moderators significant at $p < .05$. Effects were larger when a blended rather than a purely online condition was compared with face-to-face instruction; when students in the online condition were engaged in instructor-led or collaborative instruction rather than independent learning; and when the curricular materials and instruction varied between the online and face-to-face conditions. This pattern of significant moderator variables is consistent with the interpretation that the advantage of online conditions in these recent studies stems from aspects of the treatment conditions other than the use of the Internet for delivery per se.

Clark (1983) has cautioned against interpreting studies of instruction in different media as demonstrating an effect for a given medium inasmuch as conditions may vary with respect to a whole set of instructor and content variables. That caution applies well to the findings of this meta-analysis, which should not be construed as demonstrating that online learning is superior as a medium. Rather, it is the combination of elements in the treatment conditions, which are likely to include additional learning time and materials as well as additional opportunities for collaboration, that has proven effective. The meta-analysis findings do not support simply putting an existing course online, but they do support redesigning instruction to incorporate additional learning opportunities online.

Several practices and conditions associated with differential effectiveness in distance education meta-analyses (most of which included nonlearning outcomes such as satisfaction) were not found to be significant moderators of effects in this meta-analysis of Web-based online learning. Nor did tests for the incorporation of instructional elements of computer-based instruction (e.g., online practice opportunities and feedback to learners) find that these variables made a difference. Online learning conditions produced better outcomes than face-to-face learning alone, regardless of whether these instructional practices were used.

The meta-analysis did not find differences in average effect size between studies published before 2004 (which might have used less sophisticated Web-based technologies than those available since) and studies published from 2004 on (possibly reflecting the more sophisticated graphics and animations or more complex instructional designs available). Nor were differences associated with the nature of the subject matter involved.

Finally, the examination of the influence of study method variables found that effect sizes did not vary significantly with study sample size or with type of design. It is reassuring to note that, on average, online learning produced better student learning outcomes than face-to-face instruction in those studies with random-assignment experimental designs ($p < .001$) and in those studies with the largest sample sizes ($p < .01$).

The relatively small number of studies meeting criteria for inclusion in this meta-analysis limits the power of tests for moderator variables. A few contrasts that did not attain significance (e.g., time on task or type of knowledge tested) might have emerged as significant influences under a fixed-effects analysis and may prove significant when tested in future meta-analyses with a larger corpus of studies.

The narrative synthesis of studies comparing variations of online learning provides some additional insights with respect to designing effective online learning experiences. The practice with the strongest evidence of effectiveness is inclusion of mechanisms to prompt students to reflect on their level of understanding as they are learning online. In a related vein, there is some evidence that online learning environments with the capacity to individualize instruction to a learner's specific needs improves effectiveness.

As noted in chapter 4, the results of studies using purely online and blended conditions cast some doubt on the meta-analysis finding of larger effect sizes for studies blending online and face-to-face elements. The inconsistency in the implications of the two sets of studies underscores the importance of recognizing the confounding of practice variables in most studies. Studies using blended learning also tend to involve more learning time, additional instructional resources, and course elements that encourage interactions among learners. This confounding leaves open the possibility that one or all of these other practice variables, rather than the blending of online and offline media per se, accounts for the particularly positive outcomes for blended learning in the studies included in the meta-analysis.

Comparison With Meta-Analyses of Distance Learning

Because online learning has much in common with distance learning, it is useful to compare the findings of the present meta-analysis with the most comprehensive recent meta-analyses in the distance-learning field. The two most pertinent earlier works are those by Bernard et al. (2004) and Zhao et al. (2005). As noted above, the corpus in this meta-analysis differed from the earlier quantitative syntheses, not only in including more recent studies but also in excluding studies that did not involve Web-based instruction and studies that did not examine an objective student learning outcome.

Bernard et al. (2004) found advantages for asynchronous over synchronous distance education, a finding that on the surface appears incongruent with the results reported here. On closer

inspection, however, it turns out that the synchronous distance-education studies in the Bernard et al. corpus were mostly cases of a satellite classroom yoked to the main classroom where the instructor taught. It is likely that the nature of the learning experience and extent of collaborative learning were quite different in the primary and distant classrooms in these studies. For asynchronous distance education, Bernard et al. also found that the distance-education condition tended to have more favorable outcomes when opportunities for computer-mediated communication were available. Online learners in all of the studies in this meta-analysis had access to computer-mediated communication and in every case there were mechanisms for asynchronous communication.

Zhao et al. (2005) found advantages for blended learning (combining elements of online and face-to-face communication) over purely online learning experiences, a finding similar to that of this meta-analysis. Zhao et al. also found that instructor involvement was a strong mediating variable. Distance learning outcomes were less positive when instructor involvement was low (as in "canned" applications), with effects becoming more positive, up to a point, as instructor involvement increased. At the highest level of instructor involvement (which would suggest that the instructor became dominant and peer-to-peer learning was minimized), effect size started to decline in the corpus of studies Zhao et al. examined. Although a somewhat different construct was tested in the Learning Experience variable used here, the present results are consonant with those of Zhao et al. Studies in which the online learners worked with digital resources with little or no teacher guidance were coded here as "independent/active," and this category was the one learner experience category for which the advantage of online learning failed to attain statistical significance at the $p < .05$ level or better.

The relative disadvantage of independent online learning (called "active" in our conceptual model) should not be confused with automated mechanisms that encourage students to be more reflective or more actively engaged with the material they are learning on line. As noted above, a number of studies reviewed in chapter 4 found positive effects for techniques such as prompts that encourage students to assess their level of understanding or set goals for what they will learn whereas mechanisms such as guiding questions or advance organizers had mostly null results.

Implications for K–12 Education

The impetus for this meta-analysis of recent empirical studies of online learning was the need to develop research-based insights into online learning practices for K–12 students. The research team realized at the outset that a look at online learning studies in a broader set of fields would be necessary to assemble sufficient empirical research for meta-analysis. As it happened, the initial search of the literature published between 1996 and 2006 found *no* studies contrasting K–12 online learning with face-to-face instruction that met methodological quality criteria.[23] By

[23] The initial literature search identified several K–12 online studies comparing student achievement data collected from both virtual and regular schools (e.g., Cavanaugh et al. 2004; Schollie 2001), but these studies were neither experiments nor quasi-experiments with statistical control for preexisting differences between groups. Some of these K–12 studies used a pre-post, within-subject design without a comparison group; others were quasi-experiments without a statistical control for preexisting differences among study conditions (e.g., Karp and Woods 2003; Long and Stevens 2004; Stevens 1999). Several studies used experimental designs with K–12 students but did not report the data needed to compute or estimate effect sizes. A few experiments compared a K–

performing a second literature search with an expanded time frame (through July 2008), the team was able to greatly expand the corpus of studies with controlled designs and to identify five controlled studies of K–12 online learning with seven contrasts between online and face-to-face conditions. This expanded corpus still comprises a very small number of studies, especially considering the extent to which secondary schools are using online courses and the rapid growth of online instruction in K–12 education as a whole. Educators making decisions about online learning need rigorous research examining the effectiveness of online learning for different types of students and subject matter as well as studies of the relative effectiveness of different online learning practices.

---

12 online intervention with a condition in which there was no instruction (e.g., Teague and Riley 2006). Many of the references (8 out of 14) used for the Cavanaugh et al. (2004) meta-analysis of K–12 online studies were databases of raw student performance data and did not describe learning conditions, technology use or learner/instructor characteristics. A recent large-scale study by the Florida TaxWatch (2007) failed to control for preexisting differences between the students taking courses online and those taking them in conventional classrooms.

References

Reference Key

[a]Category 1 and 2 studies used in the meta-analysis
[b]Category 3 studies used in the narrative summary
[c]Category 3 studies also included in the Category 1 and 2 studies
[d]Category 3 studies reviewed but not cited in the narrative summary
[e]Additional cited references.

[a]Aberson, C. L., D. E. Berger, M. R. Healy, and V. L. Romero. 2003. Evaluation of an interactive tutorial for teaching hypothesis testing concepts. *Teaching of Psychology* 30 (1):75–78.

[a]Al-Jarf, R. S. 2004. The effects of Web-based learning on struggling EFL college writers. *Foreign Language Annals* 37 (1):49–57.

[d]Beal, T., K. J. Kemper, P. Gardiner, and C. Woods. 2006. Long-term impact of four different strategies for delivering an online curriculum about herbs and other dietary supplements. *BMC Medical Education* 6:39.

[b]Beeckman, D., L. Schoonhoven, H. Boucque, G. VanMaele, and T. Defloor. 2008. Pressure ulcers: E-learning to improve classification by nurses and nursing students. *Journal of Clinical Nursing* 17 (13):1697–1707.

[b]Beile, P. M., and D. N. Boote. 2002. Library instruction and graduate professional development: Exploring the effect of learning environments on self-efficacy and learning outcomes. *Alberta Journal of Educational Research* 48 (4):364–67.

[a]Bello, G., M. A. Pennisi, R. Maviglia, S. M. Maggiore, M. G. Bocci, L. Montini, L., and M. Antonelli. 2005. Online vs. live methods for teaching difficult airway management to anesthesiology residents. *Intensive Care Medicine* 31 (4):547–52.

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

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

[b]Bernard, R. M., and K. Lundgren-Cayrol. 2001. Computer conferencing: An environment for collaborative project-based learning in distance education. *Educational Research and Evaluation* 7 (2–3):241–61.

[a]Beyea, J. A., E. Wong, M. Bromwich, W. W. Weston, and K. Fung. 2008. Evaluation of a particle repositioning maneuver web-based teaching module. *The Laryngoscope* 118 (1):175–180.

[e]Biostat Solutions. 2006. *Comprehensive Meta-Analysis* (Version 2.2.027). [Computer software]. Mt. Airy, Md.: Biostat Solutions.

[b]Bixler, B. A. 2008. The effects of scaffolding student's problem-solving process via question prompts on problem solving and intrinsic motivation in an online learning environment. PhD diss., The Pennsylvania State University, State College, Penn.

[e]Bransford, J. D., A. L. Brown, and R. R. Cocking. 1999. *How people learn: Brain, mind, experience, and school.* Washington, D.C.: National Academy Press.

[d]Brown, B. A., and K. Ryoo. 2008. Teaching science as a language: A "content-first" approach to science teaching. *Journal of Research in Science Teaching* 45 (5):529–53.

[b, c]Caldwell, E. R. 2006. A comparative study of three instructional modalities in a computer programming course: Traditional instruction, Web-based instruction, and online instruction. PhD diss., University of North Carolina at Greensboro.

[b]Campbell, M., W. Gibson, A. Hall, D. Richards, and P. Callery. 2008. Online vs. face-to-face discussion in a Web-based research methods course for postgraduate nursing students: A quasi-experimental study. *International Journal of Nursing Studies* 45 (5):750–59.

[b]Castaneda, R. 2008. The impact of computer-based simulation within an instructional sequence on learner performance in a Web-based environment. PhD diss., Arizona State University, Tempe.

[e]Cavanaugh, C. 2001. The effectiveness of interactive distance education technologies in K–12 learning: A meta-analysis. *International Journal of Educational Telecommunications* 7 (1):73–78.

[e]Cavanaugh, C., K. J. Gillan, J. Kromrey, M. Hess, and R. Blomeyer. 2004. *The effects of distance education on K–12 student outcomes: A meta-analysis*. Naperville, Ill.: Learning Point Associates. http://www.ncrel.org/tech/distance/index.html (accessed March 5, 2009).

[b, c]Cavus, N., H. Uzonboylu, and D. Ibrahim. 2007. Assessing the success rate of students using a learning management system together with a collaborative tool in Web-based teaching of programming languages. *Journal of Educational Computing Research* 36 (3):301–21.

[e]Chang, C. C. 2008. Enhancing self-perceived effects using Web-based portfolio assessment. *Computers in Human Behavior* 24 (4):1753–71.

[b]Chang, M. M. 2007. Enhancing Web-based language learning through self-monitoring. *Journal of Computer Assisted Learning* 23 (3):187–96.

[b]Chen, B. 2007. Effects of advance organizers on learning and retention from a fully Web-based class. PhD diss., University of Central Florida, Orlando.

[e]Childs, J. M. 2001. *Digital skill training research: Preliminary guidelines for distributed learning* (Final report). Albuquerque, N. Mex.: TRW. Available through http://www.stormingmedia.us/24/2471/A247193.html (accessed March 5, 2009).

[b]Choi, I., S. M. Land, and A. J. Turgeon. 2005. Scaffolding peer-questioning strategies to facilitate metacognition during online small group discussion. *Instructional Science* 33 (5–6):483–511.

[b]Chung, S., M.-J. Chung, and C. Severance. 1999, October. Design of support tools and knowledge building in a virtual university course: Effect of reflection and self-explanation prompts. Paper presented at the WebNet 99 World Conference on the WWW and Internet Proceedings, Honolulu, Hawaii. (ERIC Document Reproduction Service No. ED448706).

[e]Clark, R. E. 1983. Reconsidering research on learning from media. *Review of Educational Research* 53 (4):445–49.

[e]Clark, R. E. 1994. Media will never influence learning. *Educational Technology Research and Development* 42 (2):21–29.

Cohen, J. 1992. A power primer. *Psychological Bulletin*, 112: 155–159.

[b]Cook, D. A., D. M. Dupras, W. G. Thompson, and V. S. Pankratz. 2005. Web-based learning in residents' continuity clinics: A randomized, controlled trial. *Academic Medicine* 80 (1):90–97.

[b]Cook, D. A., M. H. Gelula, D. M. Dupras, and A. Schwartz. 2007. Instructional methods and cognitive and learning styles in Web-based learning: Report of two randomised trials. *Medical Education* 41 (9):897–905.

[b]Crippen, K. J., and B. L. Earl. 2007. The impact of Web-based worked examples and self-explanation on performance, problem solving, and self-efficacy. *Computers & Education* 49 (3):809–21.

[b, c]Davis, J. D., M. Odell, J. Abbitt, and D. Amos. 1999, March. Developing online courses: A comparison of Web-based instruction with traditional instruction. Paper presented at the Society for Information Technology & Teacher Education International Conference, Chesapeake, Va. http://www.editlib.org/INDEX.CFM?fuseaction=Reader.ViewAbstract&paper_id=7520 (accessed March 5, 2009).

[a]Day, T. M., M. R. Raven, and M. E. Newman. 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.

[a]DeBord, K. A., M. S. Aruguete, and J. Muhlig. 2004. Are computer-assisted teaching methods effective? *Teaching of Psychology* 31 (1):65–68.

[e]Dede, C., ed. 2006. *Online professional development for teachers: Emerging models and methods.* Cambridge, Mass.: Harvard Education Publishing Group.

[b]De Wever, B., M. Van Winckel, and M. Valcke 2008. Discussing patient management online: The impact of roles on knowledge construction for students interning at the paediatric ward. *Advances in Health Sciences Education* 13 (1):25–42.

[b]Dinov, I. D., J. Sanchez, and N. Christou. 2008. Pedagogical utilization and assessment of the statistic online computational resource in introductory probability and statistics courses. *Computers & Education* 50 (1):284–300.

[d]Duphorne, P. L., and C. N. Gunawardena. 2005. The effect of three computer conferencing designs on critical thinking skills of nursing students. *American Journal of Distance Education* 19 (1):37–50.

[a]El-Deghaidy, H., and A. Nouby. 2008. Effectiveness of a blended e-learning cooperative approach in an Egyptian teacher education programme. *Computers & Education* 51 (3):988–1006.

[a]Englert, C. S., Y. Zhao, K. Dunsmore, N. Y. Collings, and K. Wolbers. 2007. Scaffolding the writing of students with disabilities through procedural facilitation: Using an Internet-based technology to improve performance. *Learning Disability Quarterly* 30 (1):9–29.

[b]Evans, K. L. 2007. Learning stoichiometry: A comparison of text and multimedia instructional formats. PhD diss., University of Pittsburgh, Penn.

[e]Florida TaxWatch. 2007. Final report: A comprehensive assessment of Florida virtual school. Tallahassee, Fla.: Florida TaxWatch Center for Educational Performance and Accountability. www.floridataxwatch.org/resources/pdf/110507FinalReportFLVS.pdf (accessed March 5, 2009).

[d]Fox, E. J., and H. J. Sullivan. 2007. Comparing strategies for teaching abstract concepts in an online tutorial. *Journal of Educational Computing Research* 37 (3):307–30.

[a]Frederickson, N., P. Reed, and V. Clifford. 2005. Evaluating Web-supported learning versus lecture-based teaching: Quantitative and qualitative perspectives. *Higher Education* 50 (4):645–64.

[b]Gaddis, B., H. Napierkowski, N. Guzman, and R. Muth. 2000, October. A comparison of collaborative learning and audience awareness in two computer-mediated writing environments. Paper presented at the National Convention of the Association for Educational Communications and Technology, Denver, Colo. (ERIC Document Reproduction Service No. ED455771).

[b]Gao, T., and J. D. Lehman. 2003. The effects of different levels of interaction on the achievement and motivational perceptions of college students in a Web-based learning environment. *Journal of Interactive Learning Research* 14 (4):367–86.

[a]Gilliver, R. S., B. Randall, and Y. M. Pok. 1998. Learning in cyberspace: Shaping the future. *Journal of Computer Assisted Learning* 14 (3):212–22.

[b]Golanics, J. D., and E. M. Nussbaum. 2008. Enhancing online collaborative argumentation through question elaboration and goal instructions. *Journal of Computer Assisted Learning* 24 (3):167–80.

[b]Grant, L. K., and M. Courtoreille. 2007. Comparison of fixed-item and response-sensitive versions of an online tutorial. *Psychological Record* 57 (2):265–72.

[d]Griffin, T. J. 2007. Liberating the spacing effect from the laboratory: A practical application in a worldwide Web-based religious education volunteer-teacher training program. PhD diss., Utah State University, Logan.

[a]Hairston, N. R. 2007. Employees' attitudes toward e-learning: Implications for policy in industry environments. PhD diss., University of Arkansas, Fayetteville.

[e]Harlen, W., and S. Doubler. 2004. Can teachers learn through enquiry online? Studying professional development in science delivered online and on-campus. *International Journal of Science Education* 26 (10):1247–67.

[a]Harris, J. M., T. E. Elliott, B. E. Davis, C. Chabal, J. V. Fulginiti, and P. G. Fine. 2007. Educating generalist physicians about chronic pain: Live experts and online education can provide durable benefits. *Pain Medicine* 9 (5):555–63.

[e]Hedges, L. V., and I. Olkin. 1985. *Statistical methods for meta-analysis.* Orlando, Fla.: Academic Press.

[b]Hilbelink, A. J. 2007. The effectiveness and user perception of 3-dimensional digital human anatomy in an online undergraduate anatomy laboratory. PhD diss. dissertation, University of South Florida, Orlando.

[e]Hiltz, S. R., and R. Goldman, eds. 2005. *Learning together online: Research on asynchronous learning networks*. Mahwah, N.J.: Lawrence Erlbaum.

[b]Hron, A., F. W. Hesse, U. Cress, and C. Giovis. 2000. Implicit and explicit dialogue structuring in virtual learning groups. *British Journal of Educational Psychology* 70 (1):53–64.

[e]Hu, P. J. H., W. Hui, T. H. K. Clark, and K. Y. Tam. 2007. Technology-assisted learning and learning style: A longitudinal field experiment. Systems, man and cybernetics, part A. *IEEE Transactions on Professional Communication 37* (6):1099 –1112.

[a]Hugenholtz, N. I. R., E. M. de Croon, P. B. Smits, F. J. H. van Dijk, and K. Nieuwenhuijsen. 2008. Effectiveness of e-learning in continuing medical education for occupational physicians. *Occupational Medicine* 58 (5):370–72.

[e]Jaffe, R., E. Moir, E. Swanson, and G. Wheeler. 2006. EMentoring for Student Success: Online mentoring and professional development for new science teachers. In *Online professional development for teachers: Emerging models and methods,* ed. C. Dede, 89–116. Cambridge, Mass.: Harvard Education Press.

[a]Jang, K. S., S. Y. Hwang, S. J. Park, Y. M. Kim, and M. J. Kim. 2005. Effects of a Web-based teaching method on undergraduate nursing students' learning of electrocardiography. *The Journal of Nursing Education.* 44 (1):35–39.

[d]Jia, J. 2007. The effects of concept mapping as advance organizers in instructional designs for distance learning programs. PhD diss., Wayne State University, Detroit, Mich..

[e]Karp, G. G., and M. L. Woods. 2003. Wellness NutriFit online learning in physical education for high school students. *The Journal of Interactive Online Learning* 2, http://www.ncolr.org/jiol/Issues/PDF/2.2.3.pdf (accessed March 5, 2009).

[b]Keefe, T. J. 2003. Using technology to enhance a course: The importance of interaction. *EDUCAUSE Quarterly* 1:24–34.

[d]Kemper, K. J., P. Gardiner, J. Gobble, A. Mitra, and C. Woods. 2006. Randomized controlled trial comparing four strategies for delivering e-curriculum to health care professions. *BMC Medical Education* 6: 2. http://www.biomedcentral.com/1472-6920/6/2 (accessed March 5, 2009).

[b]Kerfoot, B. P. 2008. Interactive spaced education versus Web-based modules for teaching urology to medical students: a randomized controlled trial. *The Journal of Urology* 179 (6):2351–57.

[d]Kerfoot, B. P., H. E. Baker, M. O. Koch, D. Connelly, D. B. Joseph, and M. L. Ritchey. 2007. Randomized, controlled trial of spaced education to urology residents in the United States and Canada. *The Journal of Urology* 177 (4):1481–87.

[d]Kerfoot, B. P., P. R. Conlin, T. Travison, and G. T. McMahon. 2007. Web-based education in systems-based practice: A randomized trial. *Archives of Internal Medicine* 167 (4):361–66.

[d]Kim, K.-H. 2006. Enhancement of secondary special education teachers' knowledge and competencies in working with families through online training modules. PhD diss., University of Kansas, Lawrence.

[d]Kock, N., and R. C. J. Chatelain-Jardon. 2008. An experimental study of simulated Web-based threats and their impact on knowledge communication effectiveness. *IEEE Transactions on Professional Communication* 51:183–97.

[d]Lawless, K. A., P. G. Schrader, and H. J. Mayall. 2007. Acquisition of information online: Knowledge, navigation and learning outcomes. *Journal of Literacy Research* 39 (3):289–306.

[d]Lee, J.-L., G. Orwig, G. A. Gunter, and L. Witta. 2006. The effect of cognitive styles upon the completion of a visually-oriented component of online instruction. PhD diss., University of Central Florida, Orlando.

[b]Lewis, B. A. 2002. The effectiveness of discussion forums in online learning. *Brazilian Review of Open and Distance Learning* 1 (1), http://www.abed.org.br/publique/cgi/cgilua.exe/sys/start.htm?infoid=16&sid=73&UserActiveTemplate=1por (accessed March 5, 2009).

[e]Lipsey, M. W., and D. B. Wilson. 2000. *Practical meta-analysis*. Vol. 49 of *Applied social research methods series.* Thousand Oaks, Calif.: Sage.

[b]Loar, R. S. 2007. The impact of a computer simulated case study on nurse practitioner students' declarative knowledge and clinical performance. PhD diss., University of Illinois at Urbana-Champaign.

[a]Long, M., and H. Jennings. 2005. *"Does it work?": The impact of technology and professional development on student achievement.* Calverton, Md.: Macro International.

[e]Long, J. D., and K. R. Stevens. 2004. Using technology to promote self-efficacy for healthy eating in adolescents. *Journal of Nursing Scholarship* 36 (2):134–39.

[a]Lowry, A. E. 2007. Effects of online versus face-to-face professional development with a team-based learning community approach on teachers' application of a new instructional practice. PhD diss., Johns Hopkins University, Baltimore, Md.

[d]Lu, R., and L. Bol. 2007. A comparison of anonymous versus identifiable e-peer review on college student writing performance and the extent of critical feedback. *Journal of Interactive Online Learning 6* (2):100–115.

[b]Maag, M. 2004. The effectiveness of an interactive multimedia learning tool on nursing students' math knowledge and self-efficacy. *Computers, Informatics, Nursing* 22 (1):26–33.

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

[a]Maki, W. S., and R. H. Maki. 2002. Multimedia comprehension skill predicts differential outcomes of Web-based and lecture courses. *Journal of Experimental Psychology: Applied* 8 (2):85–98.

[b]McKethan, R. N., M. W. Kernodle, D. Brantz, and J. Fischer. 2003. Qualitative analysis of the overhand throw by undergraduates in education using a distance learning computer program. *Perceptual and Motor Skills 97* (3 Pt. 1):979–89.

[b]McNamara, J. M., R. L. Swalm, D. J. Stearne, and T. M. Covassin. 2008. Online weight training. *Journal of Strength and Conditioning Research* 22 (4):1164–68.

[a]Mentzer, G. A., J. Cryan, and B. Teclehaimanot. 2007. A comparison of face-to-face and Web-based classrooms. *Journal of Technology and Teacher Education* 15 (2):233–46.

[a]Midmer, D., M. Kahan, and B. Marlow. 2006. Effects of a distance learning program on physicians' opioid- and benzodiazepine-prescribing skills. *The Journal of Continuing Education in the Health Professions* 26 (4):294–301.

[e]Moore, M. 1994. Administrative barriers to adoption of distance education. *The American Journal of Distance Education* 8 (3):1–4.

[b]Nelson, B. C. 2007. Exploring the use of individualized, reflective guidance in an educational multi-user virtual environment. *Journal of Science Education and Technology* 16 (1):83–97.

[b]Nguyen, F. 2007. The effect of an electronic performance support system and training as performance interventions. PhD diss., Arizona State University, Tempe.

[a]Nguyen, H. Q., D. Donesky-Cuenco, S. Wolpin, L. F. Reinke, J. O. Benditt, S. M. Paul, and V. Carrieri-Kohlman. 2008. Randomized controlled trial of an Internet-based versus face-to-face dyspnea self-management program for patients with chronic obstructive pulmonary disease: Pilot study. *Journal of Medical Internet Research* 10 (2). http://www.jmir.org/2008/2/e9/ (accessed March 5, 2009).

[a]Ocker, R. J., and G. J. Yaverbaum. 1999. Asynchronous computer-mediated communication versus face-to-face collaboration: Results on student learning, quality and satisfaction. *Group Decision and Negotiation* 8 (5):427–40.

[a]O'Dwyer, L. M., R. Carey, and G. Kleiman. 2007. A study of the effectiveness of the Louisiana Algebra I online course. *Journal of Research on Technology in Education* 39 (3):289–306.

[d]O'Leary, P. F., and T. J. Quinlan. 2007. Learner-instructor telephone interaction: Effects on satisfaction and achievement of online students. *American Journal of Distance Education 21* (3):133–43.

[a]Padalino, Y., and H. H. C. Peres. 2007. E-learning: A comparative study for knowledge apprehension among nurses. *Revista Latino-Americana de Enfermagem* 15:397–403.

[e]Parker, M. J. 1999, June 22–24. Are academic behaviors fostered in Web-based environments? Paper presented at the National Education Computing Conference, Atlantic City, N.J. (ERIC Document Reproduction Service No. ED432993).

[e]Pearson, P. D., R. E. Ferdig, R. L. Blomeyer, Jr., and J. Moran. 2005. *The effects of technology on reading performance in the middle school grades: A meta-analysis with recommendations*

for policy. Naperville, Ill.: Learning Point Associates. (ERIC Document Reproduction Service No. ED489534).

[a]Peterson, C. L., and N. Bond. 2004. Online compared to face-to-face teacher preparation for learning standards-based planning skills. *Journal of Research on Technology in Education* 36 (4):345–61.

[e]Picciano, A. G., and J. Seaman. 2007. *K–12 online learning: A survey of U.S. school district administrators*. Boston: Sloan Consortium. http://www.sloan-c.org/publications/survey/K-12_06.asp (accessed March 5, 2009).

[e]———. 2007. *K–12 online learning: A survey of U.S. school district administrators*. Boston: Sloan Consortium. http://www.sloan-c.org/publications/survey/K-12_06.asp (accessed March 5, 2009).

[b]Poirier, C. R., and R. S. Feldman. 2004. Teaching in cyberspace: Online versus traditional instruction using a waiting-list experimental design. *Teaching of Psychology* 31 (1):59–62.

[e]Riel, M., and L. Polin. 2004. Online communities: Common ground and critical differences in designing technical environments. In *Designing for virtual communities in the service of learning,* ed. S. A. Barab, R. Kling, and J. H. Gray, 16–50. Cambridge, Mass.: Cambridge University Press.

[a]Rockman et al. 2007. ED PACE final report. Submitted to the West Virginia Department of Education. San Francisco: Author. www.rockman.com/projects/146.ies.edpace/finalreport (accessed March 5, 2009).

[d]Romanov, K., and A. Nevgi. 2006. Learning outcomes in medical informatics: Comparison of a WebCT course with ordinary web site learning material. *International Journal of Medical Informatics* 75 (2):156–62.

[b]Ruchti, W. P., and M. R. Odell. 2002, February. Comparison and evaluation of online and classroom instruction in elementary science teaching methods courses. Paper presented at the 1st Northwest NOVA Cyber-Conference, Newberg, Ore.. http://nova.georgefox.edu/nwcc/arpapers/uidaho.pdf (accessed March 5, 2008).

[b]Ryan, R. 2007. The effects of Web-based social networks on student achievement and perception of collaboration at the middle school level. PhD diss., Touro University International, City, Calif.

[b]Saito, H., and K. Miwa. 2007. Construction of a learning environment supporting learners' reflection: A case of information seeking on the Web. *Computers & Education* 49 (2):214–29.

[d]Schaad, D. C., E. A. Walker, F. M. B. Wolf, M. Douglas, S. M. Thielke, and L. Oberg. 1999. Evaluating the serial migration of an existing required course to the World Wide Web. *Academic Medicine* 74 (10):84–86.

[d]Scheines, R., G. Leinhardt, J. Smith, and K. Cho. 2005. Replacing lecture with Web-based course materials. *Journal of Educational Computing Research* 32 (1):1–25.

[a]Schilling, K., J. Wiecha, D. Polineni, and S. Khalil. 2006. An interactive Web-based curriculum on evidence-based medicine: Design and effectiveness. *Family Medicine*. 38 (2):126–32.

[b, c]Schmeeckle, J. M. 2003. Online training: An evaluation of the effectiveness and efficiency of training law enforcement personnel over the Internet. *Journal of Science Education and Technology* 12 (3):205–60.

[d]Schmidt, K. 2002. Classroom action research: A case study assessing students' perceptions and learning outcomes of classroom teaching versus online teaching. *Journal of Industrial Teacher Education* 40 (1):45–59.

[b]Schnitman, I. 2007. The dynamics involved in Web-based learning environment (WLE) interface design and human-computer interactions (HCI): Connections with learning performance. PhD diss., West Virginia University, Morgantown.

[a]Schoenfeld-Tacher, R., S. McConnell, and M. Graham. 2001. Do no harm: A comparison of the effects of online vs. traditional delivery media on a science course. *Journal of Science Education and Technology* 10 (3):257–65.

[e]Schollie, B. 2001. Student achievement and performance levels in online education research study. Edmonton, Alberta: Alberta Online Consortium. http://www.albertaonline.ab.ca/pdfs/AOCresearch_full_report.pdf (accessed March 5, 2008).

[b]Schroeder, B. A. 2006. Multimedia-enhanced instruction in online learning environments. PhD diss., Boise State University, Boise, Idaho.

[b]Schutt, M. 2007. The effects of instructor immediacy in online learning environments. PhD diss., University of San Diego and San Diego State University, San Diego, Calif.

[e]Schwen, T. M., and N. Hara. 2004. Community of practice: A metaphor for online design. In *Designing for virtual communities in the service of learning,* ed. S. A. Barab, R. Kling, and J. H. Gray, 154–78. Cambridge, U.K.: Cambridge University Press.

[b]Scoville, S. A., and T. D. Buskirk. 2007. Traditional and virtual microscopy compared experimentally in a classroom setting. *Clinical Anatomy* 20 (5):565–70.

[a]Sexton, J. S., M. R. Raven, and M. E. Newman. 2002. A comparison of traditional and World Wide Web methodologies, computer anxiety, and higher order thinking skills in the inservice training of Mississippi 4-H extension agents. *Journal of Agricultural Education* 43 (3):25–36.

[e]Setzer, J. C., and L. Lewis. 2005. *Distance education courses for public elementary and secondary school students: 2002–03.* NCES No. 2005-010. Washington, D.C.: National Center for Education Statistics.

[b]Shen, P. D., T. H. Lee, and C. W. Tsai. 2007. Applying Web-enabled problem-based learning and self-regulated learning to enhance computing skills of Taiwan's vocational students: A quasi-experimental study of a short-term module. *Electronic Journal of e-Learning* 5 (2):147–56.

[b]Shih, Y. E. 2007. Dynamic language learning: Comparing mobile language learning with online language learning. PhD diss., Capella University, Minneapolis, Minn..

[e]Sitzmann, T., K. Kraiger, D. Stewart, and R. Wisher. 2006. The comparative effectiveness of Web-based and classroom instruction: A meta-analysis. *Personnel Psychology* 59:623–64.

[b]Smith, C. M. 2006. Comparison of Web-based instructional design strategies in a pain management program for nursing professional development. PhD diss., State University of New York at Buffalo.

[a]Spires, H. A., C. Mason, C. Crissman, and A. Jackson. 2001. Exploring the academic self within an electronic mail environment. *Research and Teaching in Developmental Education* 17 (2):5–14.

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

[e]Stevens, K. 1999. Two Canadian approaches to teaching biology, chemistry, mathematics and physics to senior high school students in virtual classes. Paper presented at the Australasian Science Education Research Association, Rotorua, New Zealand. (ERIC Document Reproduction Service No. ED451987).

[b]Suh, S. 2006. The effect of using guided questions and collaborative groups for complex problem solving on performance and attitude in a Web-enhanced learning environment. PhD diss., Florida State University, Tallahassee.

[a]Sun, K., Y. Lin, and C. Yu. 2008. A study on learning effect among different learning styles in a Web-based lab of science for elementary school students. *Computers & Education* 50 (4):1411–22.

[a]Suter, W. N., and M. K. Perry. 1997. Evaluation by electronic mail. Paper presented at the annual meeting of the Mid-South Educational Research Association, Memphis, Tenn.. (ERIC Document Reproduction Service No. ED415269).

[e]Tallent-Runnels, M. K., J. A. Thomas, W. Y. Lan, S. Cooper, T. C. Ahern, S. M. Shaw, and X. Liu. 2006. Teaching courses online: A review of research. *Review of Educational Research* 76:93–135.

[b]Tantrarungroj, P. 2008. Effect of embedded streaming video strategy in an online learning environment on the learning of neuroscience. PhD diss., Indiana State University, Terre Haute.

[e]Teague, G., and R. H. Riley. 2006. Does it improve high school students' ability to perform cardiopulmonary resuscitation in a simulated environment? *Resuscitation* 71 (3):352–57.

[d]Tian, L., S. Tang, W. Cao, K. Zhang, V. Li, and R. Detels. 2007. Evaluation of a Web-based intervention for improving HIV/AIDS knowledge in rural Yunnan, China. *AIDS* 21 (8):137–42.

[b]Tselios, N. K., N. M. Avouris, A. Dimitracopoulou, and S. Daskalaki. 2001. Evaluation of distance-learning environments: Impact of usability on student performance. *International Journal of Educational Telecommunications* 7 (4):355–78.

[e]Tucker, B. 2007, June. *Laboratories of reform*: *Virtual high schools and innovation in public education*. Washington, D.C.: Education Sector Reports. http://www.butlertech.org/ek_sitepath/uploadedfiles/Teen_Education/Butler_Tech_Online/Laboratories%20of%20Reform.pdf (accessed March 5, 2009).

[a]Turner, M. K., S. R. Simon, K. C. Facemyer, L. M. Newhall, and T. L. Veach. 2006. Web-based learning versus standardized patients for teaching clinical diagnosis: A randomized, controlled, crossover trial. *Teaching and Learning in Medicine* 18 (3):208–14.

[a]Urban, C. Q. 2006. The effects of using computer-based distance education for supplemental instruction compared to traditional tutorial sessions to enhance learning for students at-risk for academic difficulties. PhD diss., George Mason University, Fairfax, Va.

[e]Uzunboylu, H. 2004. The effectiveness of Web assisted English language instruction on the achievement and attitude of the students. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004,* ed. L. Cantoni & C. McLoughlin, 727–33. Chesapeake, Va.: Association for the Advancement of Computing in Education. (ERIC Document Reproduction Service No. ED490528).

[d]Vance, D., J. Dawson, V. Wadley, J. Edwards, D. Roenker, and M. Rizzo. 2007. The accelerate study: The longitudinal effect of speed of processing training on cognitive performance of older adults. *Rehabilitation Psychology* 52 (1):89–96.

[a]Vandeweerd, J.-M. E. F., J. C. Davies, G. L. Pichbeck, and J. C. Cotton. 2007. Teaching veterinary radiography by e-learning versus structured tutorial: A randomized, single-blinded controlled trial. *Journal of Veterinary Medical Education* 34 (2):160–67.

[e]Vrasidas, C., and G. V. Glass. 2004. Teacher professional development: Issues and trends. In *Online professional development for teachers,* ed. C. Vrasidas and G. V. Glass, 1–12). Greenwich, Conn.: Information Age.

[a]Wallace, P. E., and R. B. Clariana. 2000. Achievement predictors for a computer-applications module delivered online. *Journal of Information Systems Education* 11 (1/2):13–18.

[a]Wang, L. 2008. Developing and evaluating an interactive multimedia instructional tool: Learning outcomes and user experiences of optometry students. *Journal of Educational Multimedia and Hypermedia* 17 (1):43–57.

[b]Wang, K. H., T. H. Wang, W. L. Wang, and S. C. Huang. 2006. Learning styles and formative assessment strategy: Enhancing student achievement in Web-based learning. *Journal of Computer Assisted Learning* 22 (3):207–17.

[b]Weinberger, A., B. Ertl, F. Fischer, and H. Mandl. 2005. Epistemic and social scripts in computer supported collaborative learning. *Instructional Science* 33 (1):1–30. [Two studies are reported.]

[e]WestEd with Edvance Research. 2008. Evaluating online learning: Challenges and strategies for success. Washington, D.C.: U.S. Department of Education. http://evalonline.ed.gov/ (accessed March 5, 2009).

[e]What Works Clearinghouse. 2007. *Technical details of WWC-conducted computations.* Washington, D.C.: U.S. Department of Education. http://ies.ed.gov/ncee/wwc/pdf/conducted_computations.pdf (accessed March 5, 2009).

[e]Whitehouse, P. L., L. A. Breit, E. M. McCloskey, D. J. Ketelhut, and C. Dede. 2006. An overview of current findings from empirical research on online teacher professional development. In *Online professional development for teachers: Emerging models and methods,* ed. C. Dede, 13–29. Cambridge, Mass.: Harvard University Press.

[e]Wisher, R. A., and T. M. Olson. 2003. *The effectiveness of Web-based training.* Research Report No. 1802. Alexandria, Va.: U.S. Army Research Institute.

[a]Zacharia, Z. C. 2007. Comparing and combining real and virtual experimentation: An effort to enhance students' conceptual understanding of electric circuits. *Journal of Computer Assisted Learning* 23 (2):120–32.

[e]Zandberg, I., and L. Lewis. 2008. *Technology-based distance education courses for public elementary and secondary school students: 2002-03 and 2004-05.* (NCES 2008-08). Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

[b, c]Zhang, D. 2005. Interactive multimedia-based e-learning: A study of effectiveness. *American Journal of Distance Education* 19 (3):149–62.

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

[b]Zhang, K. 2004. Effects of peer-controlled or externally structured and moderated online collaboration on group problem solving processes and related individual attitudes in well-structured and ill-structured small group problem solving in a hybrid course. PhD diss., Pennsylvania State University, State College.

[e]Zhao, Y., J. Lei, B. Yan, C. Lai, and H. S. Tan. 2005. What makes the difference? A practical analysis of research on the effectiveness of distance education. *Teachers College Record* 107 (8):183684.

[e]Zirkle, C. 2003. Distance education and career and technical education: A review of the research literature. *Journal of Vocational Education Research 28* 2):161–81.

Appendix
Meta-Analysis Methodology

Terms and Processes Used in the Database Searches

In March 2007, researchers performed searches through the following four data sources:

1. *Electronic research databases.* Using a common set of keywords (see Exhibit A-1), searches were performed in *ERIC, PsycINFO, PubMed, ABI/INFORM,* and *UMI ProQuest Digital Dissertations*. In addition, to make sure that studies of online learning in teacher professional development and career technical education were included, additional sets of keywords, shown in Exhibit A-2, were used in additional searches of *ERIC* and *PsycINFO*.
2. *Recent meta-analyses and narrative syntheses.* Researchers reviewed the lists of studies included in Bernard et al. (2004), Cavanaugh et al. (2004), Childs (2001), Sitzmann et al. (2006), Tallent-Runnels et al. (2006), Wisher and Olson (2003), and Zhao et al. (2005) for possible inclusions. Additionally, for teacher professional development and career technical education, references from recent narrative research syntheses in those fields (Whitehouse et al. 2006; Zirkle 2003) were examined to identify potential studies for inclusion.
3. *Key journals.* Abstracts were manually reviewed for articles published since 2005 in *American Journal of Distance Education, Journal of Distance Education* (Canada), *Distance Education* (Australia), *International Review of Research in Distance and Open Education,* and *Journal of Asynchronous Learning Networks.* In addition, the *Journal of Technology and Teacher Education and Career* and *Technical Education Research* (formerly known as *Journal of Vocational Education Research*) were manually searched.
4. *Google Scholar searches.* To complement these targeted searches, researchers used limiting parameters and sets of keywords (available from the authors of this report) in the Google Scholar search engine.

Exhibit A-1. Terms for Initial Research Database Search

Technology and Education/ Training Terms	Study Design Terms[a]
Distance education	Control group
Distance learning	Comparison group
E-learning	Treatment group
Online education	Experimental
Online learning	
Online training	
Online course	
Virtual learning	
Virtual training	
Virtual & course	
Internet & learning	
Internet & training	
Internet & course	
Web-based learning	
Web-based instruction	
Web-based course	
Web-based training	
"Distributed learning"	

[a] All four terms were used in one query with "OR" if the database allowed.

Exhibit A-2. Terms for Additional Database Searches for Online Career Technical Education and Teacher Professional Development

Education Terms	Technology Terms	Study Design Terms
Career education	Distance	Control group
Vocational education	Distributed	Comparison group
Teacher education	E-learning	Experimental
Teacher mentoring	Internet	Randomized
Teacher professional development	Online	Treatment group
Teacher training	Virtual	
Technical education	Web-based	

Additional Sources of Articles

Exhibit A-3 lists the sources for the resulting 502 articles that went through full-text screening.

Exhibit A-3. Sources for Articles in the Full-Text Screening

	Number of Articles Identified and Passing Initial Screening
Total retained for full-text screen	**502**
Source of articles in full-text screen:	
Electronic research database searches	316
Additional database searches for teacher professional development and career technical education	6
Recent meta-analyses	171
Manual review of key journals	19
Google Scholar searches	31
Recommendations from experts	3
Overlaps	–36
Unretrievable	–8

Effect Size Extraction

Of the 176 studies passing the full-text screening, 99 were identified as having at least one contrast between online learning and face-to-face or offline learning (Category 1) or between blended learning and face-to-face/offline learning (Category 2). These studies were transferred to quantitative analysts for effect size extraction.

Numerical and statistical data contained in the studies were extracted for analysis with Comprehensive Meta-Analysis software (Biostat Solutions 2006). Data provided in the form of *t*-tests, *F*-tests, correlations, *p*-levels, and frequencies were used for this purpose.

During the data extraction phase, it became apparent that one set of studies rarely provided sufficient data for Comprehensive Meta-Analysis calculation of an effect size. Quasi-experimental studies that used hierarchical linear modeling or analysis of covariance with adjustment for pretests and other learner characteristics through covariates typically did not report some of the data elements needed to compute an effect size. For studies using hierarchical linear modeling to analyze effects, typically the regression coefficient on the treatment status variable (treatment or control), its standard error, and a *p*-value and sample sizes for the two groups were reported. For analyses of covariance, typically the adjusted means and *F*-statistic were reported along with group sample sizes. In almost all cases, the unadjusted standard deviations for the two groups were not reported and could not be computed because the pretest-posttest correlation was not provided. Following the advice of Robert Bernard, the chief meta-analysis expert on the project's Technical Working Group, analysts decided to retain these studies and to use a conservative estimate of the pretest-posttest correlation ($r = .70$) in estimating an effect size for those studies where the pretest was the same measure as the posttest and using a pretest-posttest correlation of $r = .50$ when it was not. These effect sizes were flagged in the coding as "estimated effect sizes," as were effect sizes computed from *t* tests, *F* tests, and *p* levels.

In extracting effect size data, the analysts followed a set of rules:

- The unit of analysis was the independent contrast between online condition and face-to-face condition (Category 1) or between blended condition and face-to-face condition (Category 2). Some studies reported more than one contrast, either by reporting more than one experiment or by having multiple treatment conditions (e.g., online vs. blended vs. face-to-face) in a single experiment.

- When there were multiple treatment groups or multiple control groups and the nature of the instruction in the groups did not differ considerably (e.g., two treatment groups both fell into the "blended" instruction category), then the weighted mean of the groups and pooled standard deviation were used.

- When there were multiple treatment groups or multiple control groups and the nature of the instruction in the groups did differ considerably (e.g., one treatment was purely online whereas the other treatment was blended instruction, both compared against the face-to-face condition), then analysts treated them as independent contrasts.

- In general, one learning outcome finding was extracted from each study. When multiple learning outcome data were reported (e.g., assignments, midterm and final examinations, grade point averages, grade distributions), the outcome that could be expected to be more stable and more closely aligned to the instruction was extracted (e.g., final examination scores instead of quizzes). However, in some studies, no learning outcome had obvious superiority over the others. In such cases, analysts extracted multiple contrasts from the study and calculated the weighted average of the multiple outcome scores if the outcome measures were similar (e.g., two final tests, one testing procedural skills and the other testing declarative knowledge). For example, in one study, analysts retained two outcome findings because the outcome measures were quite different (Schilling et al. 2006). One measure was a multiple-choice test, examining basic knowledge, whereas the other was a performance-based assessment, testing students' strategic and problem-solving skills in the context of ill-structured problems.

- Learning outcome findings were extracted at the individual level. Analysts did not extract group-level learning outcomes (e.g., scores for a group product). Too few group products were included in the studies to support analyses of this variable.

The review of the 99 studies for effect size calculation produced 50 independent effect sizes (27 for Category 1 and 23 for Category 2) from 45 studies; 54 studies did not report sufficient data to support effect-size calculation.

Coding of Study Features

All studies that provided enough effect size data were coded for their study features and for study quality. The top-level coding structure, incorporating refinements made after pilot testing, is shown in Exhibit A-4. (The full coding structure is available from the authors of this report.)

Twenty percent of the studies with sufficient data to compute effect size were coded by two researchers. The interrater reliability across these double-coded studies was 86.4 percent. As a result of analyzing coder disagreements, some definitions and decision rules for some codes were refined; other codes that required information missing in the vast majority of documents or that proved difficult to code reliably (e.g., indication of whether the instructor was certified or not) were eliminated. A single researcher coded the remaining studies.

Exhibit A-4. Top-level Coding Structure for the Meta-analysis

Study Feature Coding Categories
• Study type • Type of publication • Year of publication • Study author • Whether the instructor was trained in online training • Learner type • Learner age • Learner incentive for involvement in the study • Learning setting • Subject matter • Treatment duration • Dominant approach to learner control • Media features • Opportunity for face-to-face contact with the instructor • Opportunity for face-to-face contact with peers • Opportunity for *asynchronous* computer-mediated communication with the instructor • Opportunity for *asynchronous* computer-mediated communication with peers • Opportunity for *synchronous* computer-mediated communication with the instructor • Opportunity for *synchronous* computer-mediated communication with peers • Use of problem-based or project-based learning • Opportunity for practice • Opportunity for feedback • Type of media-supported pedagogy • Nature of outcome measure • Nature of knowledge assessed
Study Design Codes
• Unit of assignment to conditions • Sample size for unit of assignment • Student equivalence • Whether equivalence of groups at preintervention was described • Equivalence of prior knowledge/pretest scores • Instructor equivalence • Time-on-task equivalence • Curriculum material/instruction equivalence • Attrition equivalence • Contamination

The Department of Education's mission is to promote student achievement and preparation for global competitiveness by fostering educational excellence and ensuring equal access.

www.ed.gov