CHAPTER 21

Instruction

Lindsey E. Richland, Marcia C. Linn, and Robert A. Bjork
University of California, USA

COGNITION AND INSTRUCTION: BRIDGING LABORATORY AND CLASSROOM SETTINGS

Researchers who conduct laboratory studies of memory, reasoning, and forgetting, almost always with undergraduate students as participants, have much in common with researchers who conduct classroom studies of memory, reasoning, and cumulative understanding, usually with pre-college students as participants. Yet studies within these two settings have traditionally used distinct research methods and addressed different questions (Linn 1990; Brown 1992; Collins 1992; Shonkoff & Phillips 2000; Shavelson & Towne 2002; Bell *et al.* 2004). This chapter focuses on ways that research deriving from laboratory and classroom traditions can be mutually informative. In particular, this chapter focuses on studies that both address the cognitive mechanisms underlying learning *and* seek answers to questions of genuine educational importance. While this chapter has broad implications for practitioners, its primary goal is to encourage traditional laboratory researchers to broaden research programs to address complex, educationally relevant learning.

We begin with a consideration of conditions that have fostered the separation between laboratory and classroom research traditions. We then discuss two methodological approaches utilized in recent projects that broaden basic cognitive research on learning and increase its educational relevance. One approach consists of laboratory research designed to examine whether existing laboratory findings and principles extend to materials and retention intervals that are educationally realistic. The second approach consists of classroom studies that test whether principles of learning derived from laboratory research can upgrade instruction in actual classrooms. We conclude with recommendations for research methods that bridge the gap between the laboratory and classrooms and have the potential to address real-world educational problems often considered intractable. We specifically advocate partnerships among the varied stakeholders, including laboratory and classroom researchers, to address the pressing dilemmas facing educational policy makers today.

Handbook of Applied Cognition: Second Edition. Edited by Francis T. Durso.
Copyright © 2007 John Wiley & Sons, Ltd.

Rationale: Getting Beyond the Basic vs. Applied Distinction

Traditional psychological research on learning has made an implicit distinction between basic and applied investigations. Basic research, aimed at understanding human cognitive processes, has been associated with laboratory-based studies using simple, clearly defined materials, controlled conditions, and delays on the order of minutes or hours. Applied research, aimed at improving classroom instruction and promoting lifelong learning, has been associated with classroom-based studies using complex curriculum materials and assessments of students' understanding, retention, and transfer across retention intervals extending to months or years.

The overall distinction between basic and applied research has been challenged in recent years (Stokes 1997). This distinction can lead researchers to assume that any work conducted in a use-based setting, such as a classroom, is by definition an application of some finding. In contrast, Stokes (1997) argues that research in complex contexts can yield generalizable insights in its own right. Stokes argues that many important basic scientific discoveries have emerged from attempts to solve applied problems (e.g., Pasteur's germ theory of disease).

We suggest that the basic vs. applied distinction is unhelpful in research on learning and instruction because the boundaries have become blurred. The implication that basic research is necessary to guide and inform applied research is not consistent with numerous classroom studies that yield powerful findings (e.g., diSessa 2000; Songer *et al.* 2003). Instead, work in areas such as design-based research illustrates how generalizable principles of learning and instruction can be derived from iterative attempts to design and improve classroom instruction (e.g., Brown 1992; Cobb *et al.* 2003; Shavelson *et al.* 2003; Linn *et al.* 2004). Recent psychological studies also support this view by showing that educational research conducted in classroom settings (Anderson *et al.* 2004; Klahr & Li 2005) yields unique, valuable basic research insights. These studies can raise new questions for investigation and reveal unexplored assumptions made within laboratory studies (e.g., Brown 1992; Richland *et al.* 2004). Classroom studies document important educational variables, such as the role of everyday experience with physics, which are neglected in laboratory work. Classroom work can also reveal whether cognitive mechanisms identified to impact learning for simple stimuli and following short delays also guide learning for more complex materials and longer delays. Finally, classroom studies often reveal unintended consequences of laboratory findings when implemented in settings that are part of a complex system (Schofield 1995).

We suggest that the historic distinction between basic and applied research has fueled the separation between laboratory and classroom-based research. This has led to largely separate bodies of literature and reduced the cross-fertilization of ideas and findings across these settings. Stokes' (1997) framework of *use-based research* has inspired both researchers and the National Science Foundation to emphasize the utility of research that begins with educational questions or currently intractable instructional debates as sources for investigating general mechanisms underlying learning (Klahr & Li 2005). We suggest that this framework can help forge connections between laboratory and classroom-based research traditions.

Desirable Difficulties: Implications for Classroom Learning

Many commonly accepted findings from laboratory studies of learning have quite provocative and potentially important implications for education. This chapter focuses on one such cluster of laboratory-based research findings that have demonstrated the beneficial effects of increasing the apparent difficulty of initial learning opportunities. These are findings that have been categorized as "desirable difficulties," a term Bjork (1994, 1999) used to describe principles for designing instruction that make learning *seem* more difficult during acquisition, slowing the *apparent* rate of acquisition, but lead to increased long-term retention and transfer. These principles are largely counterintuitive, and teachers and students alike are regularly misled to believe that the rate of acquisition is an effective predictor of learning. However, laboratory-based research has demonstrated that this is unreliable and instead, greater difficulty and slow acquisition can be markers of richer encoding and longer-term retention.

Such desirable difficulties include using tests rather than presentations as learning events (e.g., Gates 1917; Glover 1989; McDaniel & Fisher 1991; Roediger & Karpicke 2005); spacing rather than massing study sessions (for reviews see Dempster 1988, 1989, 1996; Lee & Genovese 1988; Glenberg 1992); interleaving rather than blocking to-be-learned materials and tasks (see, e.g., Shea & Morgan 1979; Carlson & Yaure 1990); and varying the conditions of practice rather than keeping conditions constant and predictable (e.g., Catalano & Kleiner 1984; Homa & Cultice 1984; Reder *et al.* 1986; Mannes & Kintsch 1987). Each of these desirable difficulties has been well replicated in controlled experiments, usually with simple verbal or motor tasks and short retention intervals, and in a few cases with more complex real-world tasks, especially in the cognitive-motor domain. Whether such manipulations can enhance learning in the classroom remains, however, largely an open issue.

Recent studies have begun to bridge from the traditional laboratory studies to more educationally relevant materials and settings, and this process has raised new questions and areas for study. This body of research provides an excellent window into the processes of forging connections between laboratory and classroom studies of learning. Investigations have taken two approaches: (1) determining how far theoretical principles derived from laboratory research extend to educationally relevant curricula materials and substantial delays; and (2) determining how to design classroom interventions to enable all students to meet the goals of everyday instruction. In the sections that follow, we discuss representative research programs that take each approach to investigate learning principles within the cluster of desirable difficulties. We consider advantages and challenges inherent within each approach.

APPROACH 1: INCREASING THE EDUCATIONAL RELEVANCE OF LABORATORY STUDIES

Many foundational studies of learning have been conducted with simple materials, such as word pairs, word lists, or simple motor tasks, which may engage quite different processing than do more familiar and complex types of educational materials. These studies also typically measured retention over short time-intervals, not over the kinds of long-term intervals that are fundamental to the educational process and classroom contexts. We review several research programs that have employed the methodological strategy of conducting laboratory studies that are more educationally relevant than the classic studies. These efforts are conducted in the laboratory and build directly on prior research on cognitive mechanisms, but they extend basic explanations to learning that is more relevant to classroom settings.

Extending the Effects of Testing and Generation

Test effects have been explored by both classroom researchers and laboratory researchers, albeit in quite different ways. In laboratory research, test effects have typically been studied with respect to their effect on information retrieval. In classroom research, test effects have been studied primarily in the context of embedding alternative types of assessments into instruction.

Building on the laboratory tradition, research on the retrieval of information as a function of test vs. study trials has recently been extended to more educationally realistic materials by Roediger and Karpicke (2005). The test effect, namely, that tests are learning events in the sense that they enhance subsequent recall of the tested materials, has been demonstrated with a wide variety of materials and tasks in studies dating back at least to Gates's (1917) research on recitation. Overall, the history of laboratory research on test effects has demonstrated that the retrieval processes engaged by tests have several important effects: They retard forgetting of the retrieved material (e.g., Gates 1917; Hogan & Kintsch 1971; Bjork 1975; Whitten & Bjork 1977; Thompson *et al.* 1978; McDaniel & Mason 1985; Wheeler & Roediger 1992; Wheeler *et al.* 2003); they potentiate subsequent study trials (e.g., Izawa 1970); and they can impair the subsequent recall of information that is in competition with the retrieved information (e.g., Anderson *et al.* 1994). These studies have used simple materials (typically word lists, paired associates, or picture sets) to demonstrate such effects.

In the educational research domain, recent writings on assessment stress the importance of tests as components of the curriculum and emphasize performance assessments that can engage students not only in assessing their own understanding, but in learning about specific topics as a part of the assessment (Pellegrino *et al.* 2001; Shavelson & Towne 2002). In classroom studies, researchers have shown the advantage of inserting questions in study materials (see, e.g., Hamaker 1986). When educators design study materials with embedded questions that require student responses, they find that conceptual comprehension of the instruction is increased (Palinscar & Brown 1984; Scardamalia & Bereiter 1991; Chi 2000; Davis & Linn 2000). Importantly, these projects emphasize the role of conceptual tests that tap complex cognition such as making predictions, critiquing evidence, integrating topics, or building on prior knowledge. At the same time, the growing emphasis on accountability in schooling has increased reliance on standardized tests that often ask students to retrieve unconnected pieces of information and do not serve as learning events. A growing body of research demonstrates the advantages of using assessments that require the same cognitive activities emphasized in instruction and suggest that when tests ask for recall only, classroom instruction often relies on recall as well (Black & Wiliam 1998).

Extending Test Effects to Educational Materials

Roediger and Karpicke (2005) demonstrate the relevance of laboratory findings about the test effect's impact on retention to educationally relevant reading comprehension tasks. Participants in two studies studied short prose passages selected from the Test of English as a Foreign Language (TOEFL; Rogers 2001). They then had the opportunity to study some of these materials again and were also given a free-recall comprehension test on the other passages (without feedback). Finally, retention of the material was tested at a five-minute, two-day, or a one-week delay in a between-subject design. The results exhibit a striking interaction: In the immediate (five-minute) condition, participants recalled more in the study–study condition than they did in the study–test condition. After a delay of either two or seven days, however, participants showed greater memory for passages that had been tested rather than re-studied.

In a second experiment, Roediger and Karpicke examined the effects of repeated test opportunities as compared to more intensive study. In a between-subjects design, participants either studied a prose passage four times consecutively (SSSS), studied it three times consecutively and then were tested (SSST), or studied it once and were tested three times (STTT). Recall was then tested after five minutes or after one week. As in the prior experiment, Roediger and Karpicke found that there was a short-term benefit for re-studying the passage multiple times. After a week's delay, however, participants who were tested during the learning phase performed much better than learners who only studied the passage, and there was an additional small benefit for testing multiple times over testing once. Interestingly, students reported that the SSSS condition was least interesting, but predicted they would learn the most.

This study demonstrates that for complex prose passages, testing is a more powerful learning event than direct study over the long term, although direct study can show greater benefits in the short term. By increasing the educational validity of the materials implemented in this study as well as the test format, these researchers provide an important bridge to educational settings. The interaction between condition with retention interval obtained by Roediger and Karpicke replicates earlier laboratory findings, particularly those obtained by Hogan and Kintsch (1971), and thus provides a good example of a laboratory finding that carries over to materials and retention intervals that are educationally realistic. Recent studies have demonstrated success in incorporating tests into undergraduate psychology courses (Leeming 2002; McDaniel, 2004) indicating that the extension of this research to instruction is useful. Leeming (2002) gave 192 students short tests at the beginning of every class period in four psychology courses, and found higher course grades, higher retention, greater satisfaction, and fewer course withdrawals from participating students than from prior courses with only four total exams.

When advocating for the use of tests as learning events, a second question emerges from this research. Specifically, what is the effect on recall of incorrect materials either generated or considered during a test event? For example, if a test is in multiple-choice format and the learner considers three incorrect alternatives for every problem, will the test effect improve their false memory for these items as correct responses?

These questions are under investigation by Roediger and Marsh (2005) and McDermott (2006). In Roediger and Marsh's study, undergraduate participants who studied text materials were tested on the materials they read, as well as materials they did not study but might have known from prior experience. They were then tested on multiple-choice problems with two, four, or six alternatives and asked to answer every question, even if they had to guess. After a short delay, participants were given a cued-recall test on both the studied and unstudied materials and for materials tested earlier and not tested earlier. Participants were asked not to guess on this final test.

Roediger and Marsh found that the test effect was replicated overall, that is, there were benefits on the final cued-recall test of having been tested earlier via multiple-choice items. A closer examination of their results reveals, however, that the effect decreased linearly with the number of alternatives. The test effect was most pronounced when questions were initially tested with only two alternatives, it was less strong with four alternatives, and even less effective with six alternatives. Second, the number of false lures given as answers in the second test increased linearly with the number of alternatives, such that the fewest were given when initially tested with two alternatives and the most were given when initially tested with six alternatives. Participants had been asked not to guess, which suggests that production of the false lures reflected false beliefs that these answers were correct. Analyses revealed that participants who remembered and continued to choose an answer they incorrectly selected during the first test drove these error data.

These studies suggest that the test effect leads to increased retention when students produce correct information, but also enables students to learn material that is inaccurate, if they generate the inaccurate information in response to test questions. This raises the issue of feedback and explicit error correction.

Extending Laboratory Studies of Feedback

Research on feedback has a long tradition within studies of learning, and these recent results suggest that the use of testing as a learning event requires consideration of feedback strategies. Pashler *et al.* (2003) found, for example, that adding feedback to tests can foster subsequent correct recall even under conditions such as delaying the test which increase the likelihood of an error being made. Feedback has been demonstrated to be powerful in both laboratory and classroom studies, though there is reason to believe that optimal feedback conditions for learning may be somewhat different across settings and materials.

We discuss two research areas in which laboratory-based studies of feedback can be informative to educational practice. One such body of research has focused on the role of differences in the *timing* of feedback, specifically the amount of delay between when a learner is tested and when they receive feedback. Insights into the optimal timing for feedback are useful in organizing classroom instruction and designing technological learning environments.

Kulik and Kulik (1988) conducted a meta-analysis of experimental research on feedback timing using simple and educational tests. Interestingly, some discrepancies about the optimal feedback emerged based upon the nature of the learned materials. The meta-analysis revealed that immediate feedback tended to be more effective than longer delayed feedback when learning materials were more complex educational tests and in educational settings. This finding resonates with studies showing that when teachers return student homework and tests quickly, students learn more (Sloane & Linn 1988). In contrast, delayed feedback was more effective for simple stimuli and abstract materials in laboratory settings. Differences in the posttest materials may provide some insight into this finding. In the majority of the applied studies, the posttest materials were different from the exact items tested during the feedback training. By contrast, the majority of the tests of abstract material examined posttest scores on the identical training materials. Because educators are often more concerned with learners' ability to develop knowledge that can transfer to tests that have somewhat different features from the initial learning context, so the laboratory-based studies may be more relevant. Thus, increased delays to feedback could lead to greater retention but less flexibility in knowledge representations.

Alternatively, subtle differences in the time-scales between studies may also impact the different patterns of feedback delays on learning. For instance, the delayed feedback in the educational settings was typically given after day or week delays. In contrast, delayed feedback in laboratory settings and with simple stimuli tended to be given after each item or at the end of the test, with delays of the order of minutes or hours. The Kulik and Kulik (1988) findings might indicate that a short delay to feedback, on the order of minutes or hours, would be most optimal in a classroom setting.

Another factor may have been that only a small selection of the laboratory studies, and none of the applied studies, tested the impact of feedback timing on a delayed test. In the motor literature and cognitive tests where items are learned during testing, immediate feedback is demonstrated to be more effective than delayed feedback on a test after a minimal delay, though delayed feedback is reliably more effective after a longer test delay (see, e.g., Kulhavy 1977; Schmidt *et al.* 1989; Winstein & Schmidt 1990; Schmidt 1991). Based on these types of findings, the desirable-difficulty framework would recommend delayed feedback in order to produce longer-term retention, although short-term gains might be obtained through immediate feedback (Bjork 1994, 1999). More research is necessary to determine whether this would hold in more complex settings, or whether, as indicated in Kulik and Kulik's (1988) analysis, there are multiple determinants for the impact of feedback on longer-term retention.

Laboratory research has also been able to focus on the relationship between specific characteristics of test items and feedback. For instance, recent studies have clarified the interplay between confidence in incorrect prior knowledge and feedback on educational materials. Butterfield and Metcalfe (2001) found that errors made with high confidence were "hyper-corrected" by feedback – that is, the errors were most likely to be replaced by correct answers on a delayed test. In this study, participants were tested on their prior knowledge for general information trivia items. People were asked to rate their confidence in the accuracy of their responses to free recall questions, and then were given immediate feedback. Feedback consisted of both a statement of their accuracy and the correct response if they had been in error.

Butterfield and Metcalfe were thus able to identify high confidence errors, that is, items on which a given participant had initially indicated high confidence in their accuracy, but that in fact were errors. After a delay, participants were given a final test on a subset of the same set of trivia items, half that were answered correctly, and half that were answered incorrectly. People were asked to produce the answer they believed to be most correct and to produce two other responses that came to mind. One might expect that the high confidence errors would be resistant to change, and thus these items would be less accurate at posttest. In contrast, findings revealed that participants were more likely to correct errors for content in which they had originally expressed high confidence in their accuracy.

People did remember their initial answers, and frequently listed them within the three potential responses, but they successfully identified that these were not correct.

These findings build on early studies in which learners who answered a question incorrectly were found to study feedback more carefully if they had previously assumed the answers were correct than if they had not expected to answer the question correctly (Kulhavy & Stock 1989). Interestingly, these authors also found that these results could be increased by adding a small delay of minutes between test and feedback. This dovetails nicely with the emerging principle that a short delay to feedback may be optimal for promoting longer-term retention of some complex content.

Overall, these data reveal that feedback can play a substantial role in making testing an effective learning opportunity. In particular, these laboratory-based studies provide insights into the optimal timing of feedback, the interaction of feedback and content complexity, and the interaction between feedback and confidence ratings. Determining ideal feedback strategies can have implications for classroom teachers and for the design of technology-based curricula. The role of feedback in correcting high-confidence errors is particularly important since these are likely to be persistent sources of misconceptions within classroom learning when uncorrected. Care needs to be taken, however, in attempting to remedy strongly held beliefs because these often have experience-based justifications that deserve attention (diSessa 2000; Linn & Hsi 2000). For example, when students argue that metals are colder than wood at room temperature because they feel colder, the remedy needs to respect the tactile evidence and help learners reinterpret their experiences.

Extending the Generation Effect

A second learning principle identified in the laboratory that is closely related to the test effect is the generation effect (Slamecka & Graf 1978). In the test-effect studies, students typically answer questions that address information studied earlier in the experiment. In generation-effect studies, a similar procedure is sometimes employed, making the two effects essentially the same, or participants are asked to generate answers based on their prior knowledge, not based on recently studied information (thus, to give a simple example, participants might be asked to generate the incomplete words in "A weather phenomenon: Th**d*r and L*gh*n**g"). Generation is then compared to conditions in which learners only read or listen to the material. As in the case of test effects, materials that are generated tend to be recalled better than words that are studied, and often recalled much better (see, e.g., Jacoby 1978). Because classroom learning relies upon both acquisition of new knowledge and retrieval of prior knowledge, both laboratory-based learning principles have relevance. Generation provides one specific framework for conceptualizing *active participation* in classroom instruction, by enabling learners to retrieve prior knowledge as part of their acquisition of new learning. Engaging students in active participation has been a major part of educational reform recommendations, though the precise meaning of this term has been interpreted in many ways.

In the decades since the initial generation-effect experiments, such as the study by Slamecka and Graf, which tended to employ simple materials, the generation effect has been found to be robust, with a wide set of more educationally relevant materials. deWinstanley (1995), for example, replicated the generation effect using trivia facts and

---

Pesta *et al.* (1999) demonstrated that the generation effect improves recall of multiplication facts.

Educationally, another important question is whether students can learn the benefit of generation as a learning strategy, and thus whether its use during instruction will lead students to improve their own learning. Learning to learn has long been a primary goal of educational settings (Brown 1992). deWinstanley and Bjork (2004) investigated whether students could learn to use the generation effect to improve their own reading comprehension and whether student participants would learn to monitor and use generation methods independently when they participated in an experiment in which they could experience the benefits of generation over study. In two experiments, participants studied two paragraphs of science text which were each presented one phrase at a time. In a within-subjects manipulation, participants were either required to generate a critical word in each phrase or were asked to read the critical word. People completed a first paragraph in which they were given both read and generated phrases and were tested on that paragraph, via a fill-in-the-blank test for the critical words, before moving on to the second paragraph. The order of the paragraphs was counterbalanced across subjects.

For the first of the studied paragraphs, deWinstanley and Bjork replicated the generation effect, showing that learners retrieved items they had generated more successfully than items they had read. Surprisingly, following the second paragraph, participants performed as well on items they had read as words they had generated and at about the level of the generated items in the first paragraph. This finding, replicated in a second and third experiment, lends support for the conclusion proposed by deWinstanley and Bjork that learners discovered a more effective processing strategy by their second study opportunity. Specifically, they argue that participants observed the benefit of generation in the first paragraph and then used this technique on their own for the second paragraph even for items that were designated as *read* phrases. They also demonstrate in two studies that when learners were not allowed to compare their own read and generated performance, they did not show the same change in performance. Rather, the generation condition had an advantage over read for all paragraphs. This is a potentially important finding because it provides insight into how the design of generation-effect experiences can be used to train children more generally in academic and study skills. The generation effect has not been replicated in all study designs, but these findings may provide some insight into students' understanding of the generation effect and how they can learn to apply this strategy to their own learning.

Recent experimental studies have used students' allocations of study time to demonstrate that metacognitive awareness of desirable-difficulty principles and test difficulty can impact students' decisions about how to control their own learning (Metcalfe & Kornell 2003; Son 2004; Kornell & Metcalfe in press). These studies revealed that if students have limited study time, they monitor and control the timing for study repetitions (Son 2004) and length of study (Metcalfe & Kornell 2003; Kornell & Metcalfe in press) based upon the apparent difficulty of the learning materials in reliable, productive ways. Thus with the deWinstanley and Bjork (2004) findings, these experiments suggest that students could learn to exert control over their available study time in optimal ways through educational experiences with desirable difficulty.

In summary, studies of the test effect and the generation effect jointly shed light on the importance of asking students to respond to questions in the course of learning. These studies also provide insight into the role of desirable difficulties within instruction as a means for improving student learning and study skills. Expert learners who have a good understanding of their own learning processes may institute self-testing activities that enable them to learn more effectively. These studies suggest, however, that most learners benefit from prompts or manipulations that increase the likelihood that they generate responses and, as a result, examine their own learning.

Extending the Effects of Spacing to Educationally Realistic Retention Intervals

Another major limitation of many cognitive psychological theories for informing educational practice is that memory for learning is tested only after minutes or hours, or at most a day or two. Recent projects have begun to address whether laboratory findings extend to educationally realistic retention intervals. An illustrative set of studies by Pashler and his colleagues (Cepeda *et al.* 2006) have examined whether the spacing effect, a robust effect in laboratory studies, extends to educationally meaningful retention intervals. The spacing effect refers to the memory benefit that occurs when there is an interval between repetitions of study materials as opposed to study sessions that are consecutive, or massed. A closely related effect is the lag effect, which refers to the retention benefits of increasing the length of spacing intervals when compared with shorter spacing intervals (e.g., Tzeng 1973; Thios & D'Agostino 1976). For reviews of spacing and lag effects, see Dempster (1989, 1996).

In a quantitative meta-analysis of existing research on spacing in verbal learning paradigms, Cepeda *et al.* (2006) examined the relationship between the length of intervals between successive practice opportunities and retention. In an analysis of 317 experiments from 184 articles, they compared 958 accuracy values, 839 assessments of distributed practice, and 169 effect sizes. Overall, they found a strong positive effect of spacing over massing on long-term retention, but they also found that increasing the spacing interval beyond a certain optimal point, which is longer than the final retention interval, results in a slight decrease in long-term retention. Thus, for a given retention interval, an increase in inter-study interval causes test performance to first increase and then decrease. This meta-analysis clarifies early findings that drew attention to the power of inter-study intervals during list learning (e.g., Glenberg 1976), and highlights the importance of the length of the retention interval in decisions about the optimal timing of study.

Cepeda *et al.* (2006) noted that their meta-analysis reveals some important limitations of current laboratory research on spacing, especially that very few studies examine recall performance after delays of weeks, months, and years, that is, across educationally realistic retention intervals. They also noted that to explore applications of the spacing effect to children's learning, given educationally realistic retention intervals, it becomes necessary to understand how developmental effects interact with the effects of spacing.

As an important step in examining the spacing effect across realistic intervals, Cepeda *et al.* (2006, unpublished paper) examined foreign language, factual, and visual object learning with a substantial range of inter-study intervals and retention intervals out to six months in some cases. Again, the results suggest that some spacing vs. massing is very beneficial, but that for any given retention interval there is an optimal spacing interval and that further spacing has deleterious, if slight, effects. Cepeda *et al.* also argued that the non-monotonic relationship between the inter-study interval and retention interval takes the form that as the retention interval increases, the optimal spacing interval is a decreasing fraction of the final retention interval. These studies build on prior research demonstrating the potential for very long-term retention of educational content following spaced practice (e.g., Bahrick & Phelps 1987).

In a study of the spacing effect using mathematical materials, Rohrer and Taylor (in press) taught undergraduates how to calculate the number of permutations of a letter string in which at least one letter was repeated. Learners were then given ten practice trials, either all at once or spaced over two sessions with a one-week delay. Participants were tested one or four weeks later. Spaced practice resulted in poorer performance at the one-week interval, but better performance after the four-week delay. Additional study practice during massing made no difference – spacing was still better. This study, along with others, suggests that the spacing effect is useful not only for rote, memorized, items, but also for materials that require some generalization and application to new content features.

Studies of the spacing effect have useful implications for classroom learning. Curriculum designers and teachers make many decisions about spacing of tests as well as spacing of topics. These laboratory studies of spaced testing indicate that spaced testing of previously learned material could be quite powerful. They also imply the benefit of cumulative tests in educational settings that prompt re-study for information across a school year, or even multiple years. Unfortunately, these are not extremely common within current classroom practice, where most tests are considered final assessments of a single curriculum topic and are not repeated over spaced intervals.

The implications for curriculum organization are more nuanced. The benefits of spacing suggest a rationale for practices such as the spiral curriculum, in which a large number of topics are studied each year and then are reintroduced at regular intervals over multiple school years. Education assessments do not seem, however, to provide good support for the effectiveness of spiral curricula; indeed, some have argued that the spiral curriculum is one of the reasons why American students perform less well than their international counterparts on international comparison tests in mathematics and science (Schmidt *et al.* 2001). The realities of schooling mean that a spiral curriculum increases the total number of topics covered in a given academic year. Dramatic differences in the number of topics covered have been reported between countries that do well on international comparison tests, such as Japan and the Czech Republic, and countries that do poorly, such as the United States. For example, in Japan in middle school science, eight topics are covered, while the average for American classrooms is over 60 (Linn *et al.* 2000).

Thus, the generalization of the spacing effect to educational contexts invites new questions for study. For example, most studies of the spacing effect rely on retention of individual ideas, rather than the development of conceptual understanding. Research is needed to clarify how the spacing effect works for accumulating conceptual knowledge of topics. Nevertheless, research to date on the spacing effect clearly reveals advantages for spaced study for the long-term retention that is a hallmark of successful classroom instruction.

Overlearning in the Laboratory and Implications for the Classroom

The use of assessments as learning events raises the open question: Can there be too much of a good thing? Recent educational reforms that tout the use of standardized assessment measures to hold teachers and schools accountable for students' performance, termed "high stakes accountability," has led to increased instruction using drill practice in which learners practice on test items even after demonstrating success. Overlearning, as a laboratory procedure, consists of continued study or practice after some criterion level of mastery or recall has been achieved. Laboratory studies, some tracing back many decades, suggest that overlearning can enhance long-term retention. From an educational practice standpoint, however, two questions are relevant: (1) Do such results also obtain with educationally realistic materials and retention intervals? (2) If so, does the benefit due to overlearning justify the additional expenditure of time? While the answers to these questions are not yet known, some progress has been made toward understanding the impact of overlearning with classroom materials.

Research by Rohrer *et al.* (2005) provides some insight. In one of their experiments, designed to examine whether the benefits of overlearning are maintained over time, they had participants learn word pairs that linked cities and countries. There was a standard learning condition that consisted of five learning trials for the to-be-learned pairs, and an overlearning condition that consisted of 20 such trials. Retention tests were administered at one, three, or nine weeks. At all retention intervals participants in the overlearning condition performed better than did participants in the standard condition, but the magnitude of the difference fell substantially across those intervals. In a second experiment, the overlearning group studied the materials twice as many times as did the comparison group. On a test administered four weeks later, there were no significant differences between groups.

Caution is always necessary when interpreting the absence of differences, but these findings may well have implications for classroom practices. Specifically, they question the value of popular drill activities in mathematics and reading. From these data, it appears that students would be better off spending time learning new material rather than overlearning old material, such as math facts, through repeated testing in preparation for high-stakes assessments.

The spacing of repeated learning opportunities may also have a direct impact on the efficacy of overlearning. In a study examining the effects of a spaced repetition of previously overlearned high school mathematics content, Bahrick and Hall (1991) examined life-span retention of high school algebra and geometry concepts. These authors used cross-sectional data to compare the retention of learners who had learned the material in one time period, during high school, to the retention of learners who had restudied the same overlearned material in a later college course. They found strong relationships such that when the initial learning was spaced over several years, retention for the content remained high up to 50 years later. In contrast, when initial learning was concentrated in a single year, or a shorter time period, forgetting proceeded rapidly. The researchers found near-chance performance levels for the relevant mathematical materials when tested after the life-span retention intervals.

Overall, these studies suggest that overlearned material has the potential to remain in memory indefinitely when acquisition is spaced over a considerable interval. Without spacing repeated learning opportunities, however, overlearning may not provide a substantial benefit for long-term retention. More research is necessary to determine the optimal relationship between spacing and overlearning, but together these projects urge rethinking of the popular use of repetitive-drill instruction that is concentrated into a single time frame.

Summary

As revealed in this selective review, a growing group of cognitive psychologists has begun to extend classical studies of learning to materials and retention intervals that are educationally realistic. The initial findings suggest, in some cases, that laboratory phenomena demonstrated with simple materials and short retention intervals often *do* generalize, but in other cases there are reasons to be cautious in basing educational practices on laboratory findings obtained with materials and intervals that are unrealistic.

Overall, these studies support the claim that taking certain measures to increase the initial difficulty of a learning event can result in greater learning and retention over time. The following specific recommendations emerge from these laboratory studies of desirable difficulty using educational materials or delays:

- Studying information by reading is less effective than studying by testing. So, tests can and should be used as tools for learning and engagement as well as assessment. However, careful thought should be given to the role of false alternatives (e.g., multiple-choice questions) since these can result in increased memory for this incorrect information if not corrected. Feedback might reduce the learning for this incorrect information viewed during testing.
- Feedback is an essential part of students' learning, but the timing for when it is given is critical. In classroom settings, immediate feedback seems to be more effective than feedback after a long delay, but this may depend upon instructional goals and the length of measured delay. Specifically, a short delay (e.g., hours or a day) may be more effective than either immediate feedback after seconds or a long delay of multiple days or longer. Shorter delays (seconds up to one day) may also improve generalized knowledge acquisition, while relatively longer delays (minutes to weeks) improve memorization of precise facts. Further, feedback may have more of an impact on certain materials than for others – for instance, high confidence items that are incorrect.
- Memorization of classroom content can be improved by spacing repetitions of study rather than training on the content all at once. But it is important to ensure that the period between intervals is not so great that the prior knowledge is forgotten. Thus, for example, a spiral curriculum model of allowing an entire year or more to pass between revisits to a topic may be too long. Feedback may be important in ensuring that items are not forgotten, as well as the length of delay.
- Repeated drills on content may result in short-term memory benefits but this type of overlearning practice is not likely to improve longer-term retention. So, if memorization for content over a period of months or years was the goal, increasing intervals to days, weeks, or months between drills would be a better strategy to improve students' long-term memory.
- Learning experiences in which students see the benefits of desirable difficulties can be useful in enhancing their metacognitive sophistication and likelihood of using them to organize their own study. This has been shown in particular for the benefits of generation over more passive reading.

These are important principles with direct implications for classroom learning. Even so, the act of applying these general strategies to classroom instruction often raises new questions. It is critical, therefore, to research desirable-difficulty learning principles in the classroom as well as in the laboratory. In the next section we provide recent examples of research that raise the main issues inherent in truly bringing laboratory findings to the classroom.

APPROACH 2: TESTING LEARNING PRINCIPLES IN CLASSROOM SETTINGS

Classrooms and laboratories prompt learners to activate quite different motivational and attentional states, which can make it challenging to define cognitive principles that will generalize across settings. For this reason, laboratory researchers interested in educational learning are beginning to test cognitive principles in their context of interest: classrooms. These projects have varying research goals, including determining whether principles learned from the laboratory generalize to classroom learning, assessing whether principles derived from laboratory studies can improve classroom learning, and developing principles for the design of future curricula. Some research groups work directly with classroom teachers to develop curricula based on cognitive principles (e.g., Brown & Campione 1994; Gelman & Brenneman 2004). Other groups take advantage of technological tools to deliver instruction in regular classes (e.g., Anderson *et al.* 1995; Linn *et al.* 2004; Metcalfe *et al.* 2006, unpublished paper). Research projects in these latter two veins are reviewed to demonstrate available strategies for taking laboratory findings into classroom contexts.

Interventions Integrated into Curriculum Materials

Researchers committed to curricular reforms based on laboratory findings face a tradeoff between educational realism and research control, and often a parallel tradeoff between practitioner and professional support for the work. We describe recent examples of initiatives along this continuum, and consider both the benefits and challenges inherent in the alternative research designs.

Controlled Interventions into Classrooms

One methodology for determining whether laboratory-based learning principles generalize to more dynamic, complex classroom settings and materials is to constrain classroom instruction to create a lab-like setting. This approach can be achieved using designs in which students are pulled out for small-group testing, or where tasks are administered using a controlled technology platform that students complete individually. An example of the latter type is a study recently conducted by Metcalfe *et al.* (2006, unpublished paper) in which a computer game interface was used to embed science and English vocabulary instruction into an after-school program.

Metcalfe *et al.* designed instruction that integrated multiple learning principles consistent with the desirable-difficulty framework. They sought to demonstrate that these principles could be used together to improve long-term retention of standards-based science and English vocabulary. The study was conducted with at-risk sixth-grade students at an urban inner-city school. For five weeks, the students interacted with a game-like format in which multiple-learning principles were integrated. Among others, these included testing, generation, spaced practice, and feedback. There was also a no-study control. As predicted by the laboratory research, the game-like incorporation of learning principles did enhance students' learning and retention, after a week's delay, of the task vocabulary. The researchers also replicated these effects with English-language learners who used the technology to learn English vocabulary.

These findings provide support for the argument that the desirable-difficulties framework may have direct implications for classroom instruction. The main advantage of this study design is greater educational realism while maintaining experimental rigor and control over the execution of the instruction. This enables easily interpretable comparisons between instructional manipulations. The highly engaging task environment also has the advantage that students are likely to benefit from the difficulty manipulations, because they are only effective if a learner overcomes the challenges. The main disadvantage is that the computer program provides such support that there is still little evidence gained about the likelihood that an educator could successfully implement these principles within the myriad demands of an everyday classroom. For example, students' motivation levels in a group setting could lead to differences in how these principles operate when implemented by a teacher.

Interventions into Classroom Curricula

A second strategy for extending laboratory-based learning principles into the classroom is the introduction of interventions into classroom curricula. Such interventions require partnerships of researchers, teachers, administrators, and curriculum designers. Given these complex partnerships, creating control conditions for these studies is more difficult than it is for laboratory studies.

In a recent program of research that has made direct impact on multiple early childhood educational settings, Gelman and Brenneman (2004) describe a long-term collaboration with science teachers to develop Preschool Pathways to Science. The initiative draws on cognitive developmental theories of domain specificity, which is the theoretical argument that children's development of knowledge structures differs between content areas (see Gelman 1998). Most traditional theories of cognition and development describe domain-general processes, but a class of domain-specific learning theories (domain-specific, core knowledge and rational-constructivist) emerged in response to accumulating evidence for young children's fairly sophisticated conceptual capacities in several areas, including quantitative, physical, and biological reasoning. These domain-specific learning theories have implications for the design of learning environments. In particular, they imply the benefits of domain-relevant inputs to build on existing knowledge structures.

The Preschool Pathways to Science curriculum is a science and math program for young children in which tasks are designed in to build on domain-specific knowledge structures through selected domain-relevant inputs. Inputs included scientific vocabulary and skills with the scientific method such as observing data, predicting, and assessing predictions. The curriculum program also incorporated key advances in educational design such as constructivist learning environments and connected concepts. These led to tasks jointly developed by teachers and researchers such as making predictions, observing, and assessing predictions for "what is inside an apple."

This study is very different in form from the laboratory-based manipulations of desirable difficulty described above. As such, it raises several important issues for projects operating in classrooms settings. First, in pre-school through grade 12 educational settings, the researcher must grapple with developmental questions as well as pure information-processing considerations. Second, the content being taught is central to the learning study, as opposed to being a secondary consideration as it is during explorations of domain-general learning processes in the laboratory. Third, collaborative partnerships are extremely effective strategies for applying research findings to meet real classroom needs. Though, they also make findings difficult to separate from the specific setting (e.g., teacher skill set, administrator support, children's characteristics, etc). In addition, due to the highly collaborative, intensive nature of these partnerships, randomized experimental designs comparing interventions are often impractical or impossible.

Even so, research-based interventions through collaborations with classroom teachers are an important strategy for bridging laboratory and classroom learning. These interventions can produce generalizable principles for wide spread implementation, though researcher support may be necessary to ensure productive implementation of these principles.

Curriculum Interventions Using Technological-Enhanced Learning Environments

Collaborative interventions into classroom curriculum can also be conducted using technology-enhanced learning environments. This strategy allows for systematic tests of more general learning principles that coordinate with everyday classroom instruction. These environments deliver consistent instruction, allowing for some control over the inputs going to students, while also freeing the teacher to work individually with students in ways he or she might do ordinarily. Two examples of technology-enhanced learning environments that support this kind of research are the cognitive tutors, built by Anderson (Anderson *et al.* 1995), and the Web-Based Inquiry Science Environment (WISE, http://wise.berkeley.edu).

Anderson and his colleagues have used an architecture developed to model cognition, called Adaptive Control of Thought (ACT) theory, to design tutors that enhance student learning for procedural knowledge in the domains of algebra, geometry, and LISP programming. Procedural knowledge is emphasized over declarative knowledge, since it is assumed that inert knowledge can be learned more easily, while teaching successful knowledge manipulation and strategy use is more of an instructional challenge. These technological tools guide students to perform geometry proofs and algebra symbol manipulations successfully, as well as to consult other resources such as text and visualizations to understand these complex topics (Anderson *et al.* 1995; Koedinger & Anderson 1998; Corbett *et al.* 2001).

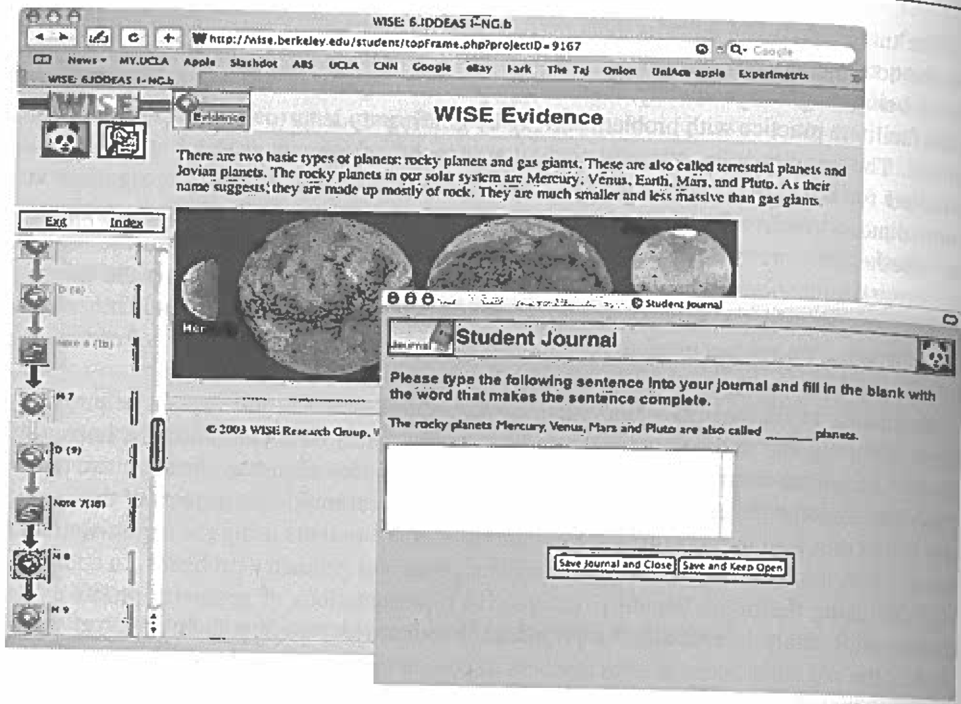
Tutors are now being designed in collaborations with curriculum experts to fit with state curriculum standards and as such have been much more fully integrated into high school classroom math and computer science instruction. These newer models also employ a knowledge-tracing technique, first implemented in the LISP tutor, in which mastery of a skill is broken into components, and students' acquisition of each component is assessed separately and required before advancement to a next instructional section.

The tutors are based on principles derived from the updated ACT-R theory, and build on the notion that instruction should be developed with reference to the cognitive competence being taught. Specifically, these tutors provide learners with some instruction, and then facilitate practice with problem solving by guiding students toward an expert solution model. The tutors invoke a "model-tracing approach," which means that a model is constructed for how an expert would solve the problem, and then learners are guided via immediate, corrective feedback to that model. The learner's response entries are evaluated for whether they are "on-path" or "off-path" actions, so generally there is a constraint that the program must be able to recognize the type of approach being used by the learner. Constraints on students' solution attempts are somewhat more minimal than in other such model-tracing tutors, but even so this approach allows for rapid diagnosing of errors and misconceptions.

Evaluation of the cognitive tutors has demonstrated their effectiveness in helping students learn algebra and geometry. In addition, researchers have examined the impact of the social context on student learning, in an attempt to understand how the cognitive tutors contribute to student success. These studies reveal several important aspects of the cognitive tutors that lead to their impact. First and foremost, students using the cognitive tutors are motivated to continue to attempt to solve algebra and geometry problems. In addition, students using the tutors benefit from creative representations of geometry proofs using means–ends analysis and algebra problems. Furthermore, as Schofield (1995) demonstrates, the cognitive tutors enable teachers to complete the text curriculum efficiently and to ensure that students who have gaps in their knowledge are able to practice the important skills necessary for them to persist in the course.

Using the cognitive-tutor technology, Anderson and his colleagues have been able to carry out well-controlled studies comparing various approaches to instruction. The cognitive tutor gathers information each second or more frequently, compared to many research studies that only gather information at the end of a day, week, or unit. Thus, researchers have an opportunity to look closely at the struggles that students face and to provide a greater understanding of educational innovations that might contribute to learning. Anderson *et al.* (2004), for example, describe eye-tracking studies that allow analysis of students' attentional process when looking at educational materials. These studies suggest that many of the ideas that govern the design of textbooks may be quite inadequate. Textbook designers often create materials designed to appeal to textbook selection committees rather than designed to improve student learning. The Anderson *et al.* (2004) eye-tracking study suggests that students' eye movements and attention are easily drawn to pictures of people and animals, and away from, for example, graphs and charts. Thus, when textbook publishers create attractive and busy pages, they may inadvertently be distracting students from the crucial information necessary for learning.

Work by researchers using the Web-based Inquiry Science Environment (WISE, http://wise.berkeley.edu) affords similar kinds of semi-controlled instruction that seamlessly integrates into classroom curriculum. WISE is an Internet-based platform that delivers science curriculum modules using inquiry-based activities. Many modules have been created through partnerships between teachers and researchers to teach standards-based science curriculum to students from approximately grades 6 through 12. Teachers can selectively identify modules that are relevant to their science instruction, and can incorporate them into their yearly curriculum. WISE provides a library of freely available modules on science content, such as astronomy, light propagation, thermal equilibrium,



WISE Evidence

There are two basic types of planets: rocky planets and gas giants. These are also called terrestrial planets and Jovian planets. The rocky planets in our solar system are Mercury, Venus, Earth, Mars, and Pluto. As their name suggests, they are made up mostly of rock. They are much smaller and less massive than gas giants.

© 2003 WISE Research Group.

Student Journal

Please type the following sentence into your journal and fill in the blank with the word that makes the sentence complete.

The rocky planets Mercury, Venus, Mars and Pluto are also called ______ planets.

Figure 21.1 Sample screens of WISE software. Information presentation and embedded prompts

and chemical reactions. These modules are also customizable, so a teacher could alter any module to best fit their teaching needs. The customizability also makes them feasible for comparative research, and embedded assessments allow for many sources of data about students' learning process.

Classroom observations of students using WISE projects reveal similarly effective benefits to those found with the cognitive tutors. Students working on science projects guided by the WISE technology tend to spend more time writing notes and conducting experiments than they did with traditional instructional materials (Linn & Hsi 2000). In addition, studies comparing different forms of animations, alternative forms of prompts, and varied discussion tools help clarify the factors that contribute to effective learning (Linn *et al.* 2004).

WISE allows designers, for example, to embed specific generation questions within the activities (see Figure 21.1). Research by Davis (1998) shows that the type of questions selected for generation can impact student learning. In Davis's work, generation questions varied along the dimension of specificity. One set of generic questions asks students to reflect on what they have learned and identify gaps in their understanding. Another set of more specific questions asks students about specific links that they might have made among materials that they had studied. Davis found that the generic questions, which ask students to self-diagnose the gaps in their understanding, were more successful than the specific questions, which students often found somewhat confusing to interpret or found too easy, and therefore said, "Well, I already know that." The Davis results resonate with the deWinstanely and Bjork study (2004), suggesting that allowing learners to assess their own learning through experience can be an effective strategy to improve their metacognitive awareness and learning skills.

Technological learning environments allow researchers to investigate the cognition underlying learning within classroom curricula and provide some measure of control over the instructional inputs. Even so, there are often limits on the generalizability of these findings to other instruction. The results of these studies often serve to improve the technological intervention itself, rather than seek to provide generalizable knowledge for use in designing alternative learning environments.

Design-Based Research

A field of design-based research has emerged to meet the goal of developing generalizable knowledge about the processes of designing effective classroom curriculum interventions. Researchers have begun to study the design process and construct databases of principles that guide and impact designers (e.g., Barab & Squire 2004; Kali *et al.* 2005). Cognition is conceptualized as being closely tied to the learning context. This research, therefore, moves away from the assumption that lies at the heart of laboratory research traditions, namely that learning principles derived in one setting can be necessarily applied to a new setting.

This body of work derives from the notion of "design experiments," as initially framed by Collins (1992) and Brown (1992). They described a process of doing experimental research that was situated within classrooms as a means for developing a generalizable body of knowledge about how cognition could be best enhanced in classroom settings. Design experiments (e.g., Brown 1992) are educational interventions that seek to investigate the basic processes of learning and instruction by manipulating classroom contexts in systematic ways. These interventions reveal the complex interplay between classroom curricula, roles of students and teachers, and assessment, all of which Brown argued must be understood and manipulated to fully characterize learning in educational settings. The nature of the interrelationships between these factors made clear experimental manipulations impossible, and involved a tradeoff between experimenter control and realism in data.

This framework has developed into a field of design-based research, in which researchers identify ways to conduct classroom research that leads to optimized learning as well as strategies to guide future designers (e.g., Cobb *et al.* 2003; Shavelson *et al.* 2003; Shavelson *et al.* 2003; Linn *et al.* 2004) characterize design studies as having several common features, though the studies themselves vary widely. They argue that design studies are typically *interventionist* and *theory-driven* in that they test theory by modifying everyday instructional activities, *iterative*, such that they contain successive modified replications of the interventions, and *utility-oriented*, in that they are concerned with producing benefits for classroom instruction. They are also *process-focused*, such that they are concerned with tracing the evolution of student and/or institutional beliefs in general as well as in response to the intervention, *multi-leveled* in developing links between classroom instruction and broader school or district structures, and *collaborative* between researchers and these various educational partners. These studies also differ from more traditional experimental studies in that multiple types of data are collected. Generally, researchers collect a detailed record of the entire study, which may include ethnographic and interview data, design process data, and evidence of student engagement and learning throughout. These data often result in a more complex picture of learning than traditional posttest performance data reveal.

Overall, this research approach is provocative and provides an alternative strategy for conceptualizing educational research. Interestingly, few of the design principles highlighted in this work have focused on learning principles deriving from laboratory research on cognition. Rather, research in this field has tended to focus on principles developed from classroom-situated learning (Kali 2006). While important and revealing, this reliance may also reflect missed opportunities to apply the rich body of laboratory-based research on learning principles, including the benefits of desirable difficulties.

Parallel Studies in Classroom and Laboratory Contexts

In our own collaborative research, we have sought to assess the relevance and applicability of certain laboratory-based learning principles to classroom curriculum through combined laboratory and classroom studies. Our partnership of cognitive psychologists, educational researchers, classroom teachers, policymakers, technology experts, and discipline experts has led to the design of studies in laboratory and classroom contexts that extend findings from laboratory studies using more educationally relevant materials. This research has been conducted with the aim of developing design principles for generalization to other instructional interventions.

Conducting parallel studies in the laboratory and the classroom allows us to test both whether factors identified as important in the laboratory impact classroom learning, and to identify factors explaining performance in the classroom. Our goal is to examine whether the principles within the desirable-difficulties framework (Bjork 1994, 1999), such as generation rather than reading, spacing rather than massing, and interleaving rather than blocking, have the potential to improve instruction. From a research methodology perspective, we have integrated laboratory and classroom studies to investigate further the value of desirable difficulties in educationally realistic learning contexts. Findings from this project so far indicate that these benefits extend to educationally realistic materials and retention intervals.

More specifically, our approach was to examine whether incorporating desirable difficulties into the design of WISE science learning modules could increase their effectiveness. Our initial goal was to examine, under controlled conditions and using introductory psychology students as participants, whether certain difficulties remained desirable when introduced into the learning of educationally realistic materials. In carefully controlled laboratory studies at UCLA, we have focused on three desirable difficulties: interleaving rather than blocking materials to be learned; having learners generate rather than re-read material presented earlier on in the study phase, and spacing rather than massing practice. Experiments have explored the relevance of these principles to science educational content. We have also used parallel materials to test these same desirable difficulties in classroom learning settings. Through the use of the WISE platform, we were able to conduct very comparable studies in both the laboratory and middle school science classrooms. This strategy allowed us to map closely between laboratory findings, which could be carefully controlled, and classroom findings, which provide insight into the generalizability of these principles from the laboratory to real educational instruction.

In one such experiment we adapted an existing WISE module on astronomy, one that covers the characteristics of planets that are important for the existence of life and is relevant to middle and high school science curriculum standards. The module covered two main characteristics: the mass of a planet, and a planet's distance from its sun. The effects of generation and interleaving, two desirable difficulties, and their interaction, were tested via a 2 × 2 design, which resulted in four conditions: Interleaving plus Generation, Interleaving plus Reading, Blocking plus Generation, and Blocking plus Reading. Interleaving was manipulated by varying whether the instruction described all of the information about the role of the Mass of a planet and then all of the information about the Distance of a planet (Blocked), or whether the instruction alternated randomly between these two sets of information (Interleaved). Generation was manipulated by varying the study opportunities that learners were given. After learning new information, they were either given a review sentence to copy into their notes (read condition) or they were asked to generate a word to fill a blank within the review sentence (generate).

The results we obtained (from a total of 96 participants) largely replicated the effects of generation and interleaving which have been obtained in laboratory studies using simpler materials. Learners who generated during study opportunities recalled significantly more of the material than learners who had simply re-read and copied the reviewed information. Interleaving materials led to greater ability to integrate information (a main goal of science education) than blocking materials, although there was little impact of interleaving on recall for facts taught during the instruction. These findings supported the hypothesis that these principles could impact complex science curriculum content learning, though the benefits of interleaving were not as large or consistent as the benefits of generation. Even so, the generation manipulation was not as challenging or active as what is advocated by science education curriculum designers.

In a second study using the interleaved version of the WISE astronomy module, the generation effect was explored using more educationally important, complex generation questions. In a between-subjects design, participants were given generation prompts that required either free response answers that integrated multiple pieces of information from either Mass OR Distance content (Single-Topic), or free response answers that integrated information from Mass AND Distance content (Topic-Integration). The latter is a more complex knowledge-integration type of reasoning, and mirrors science education pedagogical goals for students' thinking. Performance was tested on study questions, retrieval prompts, during instruction and on new questions after a two-day delay. Data from 55 undergraduates revealed that the Topic-Integration generation was more difficult and resulted in lower performance during learning but led to higher performance on new questions on a posttest following a two-day delay (Richland *et al.* 2005). This suggests that complex generation led to learning above and beyond only retention of the information generated successfully.

While these findings supported the extension of generation and interleaving principles to classroom-relevant technology environments and materials, a further step was necessary to determine whether they would impact learning in a classroom context. As argued by design-based researchers, the close relationship between cognition and context could make these principles unlikely to impact science curriculum learning within a classroom setting. Thus, a slightly modified version of the same WISE module was tested in eighth-grade classrooms in a California Bay Area public middle school. The experiment involved 140 students. As in the laboratory experiments, Interleaving and Generation were manipulated between-subjects in a 2 × 2 design. Generation was manipulated between simple generation (single fact, as in the first WISE experiment described above) and complex Intra-category generation (as in the second WISE Experiment described above). Importantly, findings revealed main effects of both generation and interleaving, indicating that these desirable difficulties have benefits that extend into the dynamic, less controlled, classroom context (Cheng 2004). When tested on new questions on a posttest, students who performed complex, inter-category generation scored higher than students who performed single fact generation. Similarly, students who were taught through interleaved material scored higher on posttest problems than students who were taught through blocked materials.

Interpreting Laboratory and Classroom Findings

Conducting parallel studies in laboratory and classroom settings has revealed dilemmas that signal underlying questions about the nature of learning. They have also led to promising research designs. If the goal of classroom learning is to have students hold up new ideas to existing views, sort out promising perspectives, link new information to related information, and organize ideas, then experimental studies of the recall of a single isolated idea may obscure the complexities of the process. The desirable difficulties that we set out to investigate, while distinct in prior laboratory settings, are more difficult to distinguish in classroom instruction. For example, the test effect and the generation effect have many similarities in complex learning of science. Testing of individuals on complex topics requires asking them to generate ideas that frequently incorporate their personal prior scientific knowledge. Yet generation is a more complex task than distinguishing whether or not a term is the correct match to a stimulus. As a result, the test effect and the generation effect become conflated when valid assessments are included in instruction.

More importantly, spacing and interleaving are conflated in complex learning because successful understanding of the topic requires making sense of related ideas, not just overcoming interference or forgetting. When teachers space instruction of specific topics, they introduce other course-related activities in between. Most often, this will include learning about alternative content within the same domain that has connections to the spaced topic (e.g., a second science topic may be taught during spaced intervals of a science lesson). When students organize their own study they may connect the spaced topic to material they encounter between repetitions. In the process of making sense of the spaced topic, students inevitably consider connections with topics and experiences that they encounter within the same classroom context. By contrast in laboratory studies, spacing and interleaving are differentiated by the degree of interference that the interleaved activity introduces. Interleaved topics are intended to slightly interfere with the learning of each individual topic. Spacing manipulations ideally do not involve active interference so much as opportunities for forgetting.

The complex interconnectedness of the cognitive processes underlying classroom instruction has traditionally made it difficult to extrapolate learning principles developed from the laboratory directly to the design of classroom instruction. Research conducted in the laboratory tradition typically focuses on single, isolated manipulations such as spacing (Bahrick & Phelps 1987), generation (Slamecka & Graf 1978), or testing (Gates 1917). In contrast, classroom interventions tend to require attention to multiple, connected manipulations. For this reason, design-based researchers have critiqued findings from laboratory studies as unclear as far as their applicability to the widely interconnected classroom setting, whereas laboratory researchers have critiqued design-based research findings as non-diagnostic as to what factor, or combination of factors, may be responsible for observed effects.

The types of studies reviewed in this chapter are increasingly filling this bi-directional gap. Laboratory researchers are extending their studies to more complex, realistic tasks and materials. Design-based researchers are using more carefully controlled classroom curriculum interventions and technology-based environments to determine how laboratory results can inform classroom practice. This body of research is growing and has the potential to provide insights into future strategies to build upon and extend laboratory research.

Overall, as this discussion suggests, desirable difficulties may have important implications for classroom learning. Understanding how they work together and how they contribute to the development of an integrated and cohesive view of a particular domain of knowledge requires investigations and research studies that go beyond examining desirable difficulties in isolation and in laboratory settings. Because implementing desirable difficulties is often counterintuitive, they are frequently neglected in classroom instruction. Teachers and students alike are regularly misled by the impression that ease of acquisition indicates successful learning. Laboratory data demonstrates that student views of their own learning often directly contradict their level of retention (Bjork 1994, 1999; Simon & Bjork 2001). Thus, this is an area in which laboratory principles of learning can have real benefits for classroom instructional design.

CONCLUSIONS

As efforts to bridge laboratory and classroom contexts in order to understand learning suggest, these activities are both important and complex. On the one hand, laboratory studies provide clear indications of specific learning principles that work reliably in laboratory studies. As research reported here suggest, however, these clear learning principles must be examined in more complex settings and with more educationally relevant materials before they can be easily applied to a classroom learning environment. The difficulties stem not just from the challenges of the materials, but also because the mechanisms that determine the clarity of these principles rely on the lack of interference or connections among the materials that are typically used in these studies.

When researchers attempt to generalize these learning principles, verified in the laboratory, to classroom contexts, they must combine them with understanding of how people learn more interconnected ideas, such as those of mathematics and science domains. These interconnections are central to extending laboratory findings to classroom settings and crucial in ensuring that individuals who are trained in our schools can engage in lifelong learning (Linn 1995; Linn & Hsi 2000; Linn *et al.* 2004). The process of lifelong learning depends on regularly revisiting ideas and making sense of their connections. Lifelong learning involves more than recalling isolated information and really rests on the importance of making analogies and inferences about when to use information, when to combine information, and when to distrust information. Learners in classroom settings need to build a more and more integrated understanding of their ideas if they are to become productive contributors to society and to lead fruitful lives.

Furthermore, studies of learning in complex settings offer an opportunity to connect learning and design. Learning principles emerging from laboratory studies can lead to new ideas that apply to the design of instructional materials. Criteria for the design of new instruction are increasingly being captured as *design principles* (e.g., Linn *et al.* 2004; Quintana *et al.* 2004; Kali 2006). These studies raise the challenge of linking learning principles and design principles to inform educational interventions. Today, researchers are beginning to connect research in complex settings, features of instruction, and design principles (Quintana *et al.* 2004; Kali 2006). Design principles capture the results of effective designs in guidelines that can be used by future designers and combined in *design patterns* (Linn & Eylon in press). Design patterns describe sequences of activities that have the potential of promoting effective understanding. These sequences, such as *predict, observe, explain*, might be a way to combine the learning principles emanating from laboratory studies with research findings when these principles are tested in classroom settings.

Efforts to merge research conducted in laboratories and research conducted in classrooms meets the goals of use-based research, identified by Stokes (1997). Several features of successful collaborations bridging laboratory and classroom contexts have emerged. First, these collaborations typically involve a partnership of researchers with expertise in laboratory and classroom learning. This leads to productive cross-fertilization of ideas, theoretical traditions, and research strategies.

Second, these ventures typically conduct classroom studies that compare educationally viable alternatives to applying the findings from the laboratory study conducted with more complex materials. Classroom studies must meet stringent criteria, such as not entailing a risk that an intervention will impede student learning. As a result, some comparisons that might be fruitful in the laboratory, such as ones designed to demonstrate the ineffectiveness of a given manipulation, are inappropriate in the classroom. An important virtue of partnerships between laboratory researchers and educational scientists is that manipulations that are at risk of being ineffective, as well as effective, can be tested in the laboratory before being considered for the classroom. In short, bridging laboratory and classroom contexts to create a science of learning offers both daunting challenges and exciting opportunities – for improving student learning, teacher learning, and school effectiveness. We are just beginning on this important trajectory.

In summary, this chapter constitutes an argument that there are compelling theoretical and practical reasons for carrying out research that bridges multiple contexts. To remedy long-standing disconnections between laboratory and classroom research, we argue for partnerships between cognitive scientists and educational researchers and we advocate an interplay of studies that incorporate educationally relevant materials and delays into laboratory studies and test findings from laboratory studies in classroom settings.

Research on cognition and instruction is *both* timely and important. It tackles problems and opportunities that characterize our educational system, and addresses the need for lifelong learning in a world that is ever more complex and rapidly changing. It builds on a foundation of informative laboratory and classroom research. And, most importantly, it responds to an eagerness among cognitive and educational scientists to bridge the laboratory and classroom settings.

AUTHOR NOTES

Preparation of this chapter was supported by Grants 011237 and 030141 from the Institute of Educational Sciences and by Collaborative Activity Grant 014448 from the James S. McDonnell Foundation. This material is also based upon work supported by the National Science Foundation under grants Nos. 9873180, 9805420, 0087832, and 0334199. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, Institute of Education Sciences, or James S. McDonnell Foundation. The authors gratefully acknowledge helpful discussions of these ideas with members of the Web-based Inquiry Science Environment group and the Technology Enhanced Learning in Science center. The authors appreciate help in production of this paper from David Crowell and Jonathan Breitbart, and we thank Jason Finley for his contributions to the experimental work we report.

REFERENCES

Anderson, M. C., Bjork, R. A. & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1063–1087.

Anderson, J. R., Corbett, A. T., Koedinger, K. & Pelletier, R. (1995). Cognitive tutors: lessons learned. *The Journal of Learning Sciences*, *4*, 167–207.

Anderson, J. R., Douglass, S. & Qin, Y. (2004). How should a theory of learning and cognition inform instruction? In A. Healy (ed.), *Experimental Cognitive Psychology and Its Applications*. Washington, DC: American Psychological Association.

Bahrick, H. P. & Hall, L. K. (1991). Lifetime maintenance of high school mathematics content. *Journal of Experiment Psychology: General*, *120*, 20–33.

Bahrick, H. P. & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *13*, 344–349.

Barab, S. A. & Squire, K. D. (2004). Design-based research: putting a stake in the ground. *Journal of the Learning Sciences*, *13*, 1–14.

Bell, P., Hoadley, C. M. & Linn, M. C. (2004). Design-based research in education. M. C. Linn, E. A. Davis & P. Bell (eds.), *Internet Environments for Science Education* (pp. 73–88). Mahwah, NJ: Lawrence Erlbaum Associates.

Bjork, R. A. (1975). Retrieval as a memory modifier: an interpretation of negative recency and related phenomena. In R. L. Solso (ed.), *Information Processing and Cognition* (pp. 123–144). New York: John Wiley & Sons.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (eds.), *Metacognition: Knowing about Knowing*. (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, R. A. (1999). Assessing our own competence: heuristics and illusions. In D. Gopher and A. Koriat (eds.), *Attention and Performance XVII. Cognitive Regulation of Performance: Interaction of Theory and Application* (pp. 435–459). Cambridge, MA: MIT Press.

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice*, *5*(1), 7–74.

Brown, A. L. (1992). Design experiments: theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, 2(2), 141–178.

Brown, A. L. & Campione, J. C. (1994). Guided discovery in a community of learners. In K. McGilly (ed.), *Classroom Lessons: Integrating Cognitive Theory and Classroom Practice* (pp. 229–270). Cambridge, MA: MIT Press.

Butterfield, B. & Metcalfe, J. (2001). Errors made with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1491–1494.

Carlson, R. A. & Yaure, R. G. (1990). Practice schedules and the use of component skills in problem solving. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *16*, 484–496.

Catalano, J. F. & Kleiner, B. M. (1984). Distant transfer in coincident timing as a function of variability of practice. *Perceptual & Motor Skills*, *58*, 851–856.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychological Bulletin*, *132*, 354–380.

Cheng, B. (2004). IDDEAS: classroom studies of desirable difficulties implemented in astronomy curricula. Paper presented at the American Education Research Association Conference. San Diego, CA.

Chi, M. T. H. (2000). Self-explaining expository tests: the dual process of generating inferences and repairing mental models. In R. Glaser (ed.), *Advances in Instructional Psychology* (Vol. 5, pp. 161–238). Mahwah, NJ: Lawrence Erlbaum Associates.

Cobb, P., diSessa, A., Lehrer, R. & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*, 9–13.

Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O. Shea (eds.), *New Directions in Educational Technology* (pp. 15–22). New York: Springer-Verlag.

Corbett, A. T., Koedinger, K. R. & Hadley, W. H. (2001). Cognitive tutors: from the research classroom to all classrooms. In P. S. Goodman (ed.), *Technology Enhanced Learning: Opportunities for Change* (pp. 235–263). Mahwah, NJ: Lawrence Erlbaum Associates.

Davis, E. A. (1998). Scaffolding students' reflection for science learning. Unpublished doctoral dissertation, University of California at Berkeley, CA.

Davis, E. A. & Linn, M. C. (2000). Scaffolding students' knowledge integration: prompts for reflection in KIE. *International Journal of Science Education*, *22*, 819–837.

Dempster, F. N. (1988). The spacing effect: a case study in the failure to apply the results of psychological research. *American Psychologist*, *43*, 627–634.

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, *1*, 309–330.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (eds.), *Memory* (Vol. 10, E. C. Carterette & M. P. Friedman, eds.), *Handbook of Perception and Cognition*. New York: Academic Press.

deWinstanley, P. A. (1995). A generation effect can be found during naturalistic learning. *Psychonomic Bulletin & Review*, *2*, 538–541.

deWinstanley, P. A. & Bjork, E. L. (2004). Processing strategies and the generation effect: implications for making a better reader. *Memory and Cognition*, *32*, 945–955.

diSessa, A. A. (2000). *Changing Minds: Computers, Learning, and Literacy*. Cambridge, MA: MIT Press.

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, *6*(40), 104.

Gelman, R. (1998). Domain specificity in cognitive development: universals and nonuniversals. In M. Sabourin, F. Craik & M. Robert (eds.), *Advances in Psychological Science, Vol. 2, Biological and Cognitive Aspects*. Hove: Psychology Press.

Gelman, R. & Brenneman, K. (2004). Science learning pathways for young children. *Early Childhood Research Quarterly*, *19*, 150–158.

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, *15*, 1–16.

Glenberg, A. M. (1992). Distributed practice effects. In L. R. Squire (ed.), *Encyclopedia of Learning and Memory* (pp. 138–142). New York: Macmillan.

Glover, J. A. (1989). The "testing" phenomenon: not gone but nearly forgotten. *Journal of Educational Psychology*, *81*, 392–399.

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research*, *56*, 212–242.

Hogan, R. M. & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, *10*, 562–567.

Homa, D. & Cultice, J. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and untilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 83–94.

Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects in paired-associate learning. *Journal of Experimental Psychology*, *83*, 340–344.

Jacoby, L. L. (1978). On interpreting the effects of repetition: solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*, 649–667.

Kali, Y. (2006). Collaborative knowledge-building using the Design Principles Database. *International Journal of Computer Support for Collaborative Learning*, *1*(2), 187–201.

Kali, Y., Spitulnik, M. & Linn, M. (2005). Design principles for educational software. Retrieved January 10, 2005, Available at: http://www.design-principles.org/dp/index.php

Klahr, D. & Li, J. (2005). Cognitive research and elementary science instruction: from the laboratory, to the classroom, and back. *Journal of Science Education and Technology*, *14*, 217–238.

Koedinger, K. R. & Anderson, J. R. (1998). Illustrating principled design: the early evolution of a cognitive tutor for algebra symbolization. *Interactive Learning Environments*, *5*, 161–180.

Kornell, N. & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *32*, 609–622.

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, *58*(1), 79–97.

Kulhavy, R. W. & Stock, W. A. (1989). Feedback in written instruction: the place of response certitude. *Educational Psychology Review*, *1*(4), 279–308.

Kulik, J. A. & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*, 79–97.

Lee, T. D. & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport*, *59*, 277–287.

Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology*, *29*, 210–212.

Linn, M. C. (1990). Establishing a science and engineering base for science education. In M. Gardner, J. G. Greeno, F. Reif, A. H. Schoenfeld, A. diSessa & E. Stage (eds.), *Toward a Scientific Practice of Science Education* (pp. 323–341). Hillsdale, NJ: Lawrence Erlbaum Associates.

Linn, M. C. (1995). Designing computer learning environments for engineering and computer science: the Scaffolded Knowledge Integration framework. *Journal of Science Education and Technology*, *4*, 103–126.

Linn, M. C., Davis, E. A. & Bell, P. (2004). Inquiry and technology. In M. C. Linn, E. A. Davis & P. Bell (eds.), *Internet Environments for Science Education* (pp. 3–28). Mahwah, NJ: Lawrence Erlbaum Associates.

Linn, M. C. & Eylon, B.-S. (in press). Science education: integrating views of learning and instruction. In P. A. Alexander & P. H. Winne (eds.), *Handbook of Educational Psychology* (2nd edn). Mahwah, NJ: Lawrence Erlbaum Associates.

Linn, M. C. & Hsi, S. (2000). *Computers, Teachers, and Peers: Science Learning Partners*. Mahwah, NJ: Lawrence Erlbaum Associates.

Linn, M. C., Lewis, C., Tsuchida, I. & Songer, N. B. (2000). Science lessons and beyond: why do U.S. and Japanese students diverge? *Educational Researcher*, *29*(3), 4–14.

McDaniel, M. A. & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, *16*, 192–201.

McDaniel, T. R. (1994). College classrooms of the future: megatrends to paradigm shifts. *College Teaching*, *42*(1), 27–31.

McDaniel, M. A. & Mason, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 371–385.

McDermott, K. B. (2006). Paradoxical effects of testing: repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory & Cognition*, *34*, 261–267.

Mannes, S. M. & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and Instruction*, *4*, 91–115.

Metcalfe, J. & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General*, *132*, 530–542.

Metcalfe, J., Kornell, N. & Son, L. K. (2006). A cognitive-science based program to enhance study efficacy in a high and low-risk setting. Submitted for publication.

Palinscar, A. S. & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*, 117–175.

Pashler, H., Zarow, G. & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory & Cognition*, *29*, 1051–1057.

Pellegrino, J. W., Chudowsky, N. & Glaser, R. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.

Pesta, B. J., Sanders, R. E. & Murphy, M. D. (1999). A beautiful day in the neighborhood: what factors determine the generation effect for simple multiplication problems? *Memory and Cognition*, *27*, 106–115.

Quintana, C., Reiser, B. J., Davis, E. A. *et al.* (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences*, *13*(3), 337–386.

Reder, L. M., Charney, D. H. & Morgan, K. I. (1986). The role of elaborations in learning a skill from an instructional text. *Memory and Cognition*, *14*, 64–78.

Richland, L. E., Bjork, R. A., Finley, J. R. & Linn, M. C. (2005). Linking cognitive science to education: generation and interleaving effects. In B. G. Bara, L. Barsalou & M. Bucciarelli (eds.), *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Richland, L. E., Holyoak, K. J. & Stigler, J. W. (2004). Analogy use in eight-grade mathematics classrooms. *Cognition and Instruction*, *22*, 37–60.

Roediger, H. L. & Karpicke, J. D. (2006). Test-enhanced learning: taking memory tests improves long-term retention, *Psychological Science*, *17*(3), 249–255.

Roediger, H. L. & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*, 1155–1159.

Rogers, B. (2001). *TOEFL CBT Success*. Lawrenceville, NJ: Peterson's.

Rohrer, D. & Taylor, K. (in press). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology*.

Rohrer, D., Taylor, K., Pashler, H. *et al.* (2005). The effect of overlearning on long-term retention. *Applied Cognitive Psychology*, *19*, 361–374.

Scardamalia, M. & Bereiter, C. (1991). Higher levels of agency for children in knowledge-building: a challenge for the design of new knowledge media. *Journal of the Learning Sciences*, *1*, 37–68.

Schmidt, R. A. (1991). Frequent augmented feedback can degrade learning: evidence and interpretations. In G. E. Stelmach & J. Requin (eds.), *Tutorials in Motor Neuroscience* (pp. 59–75). Dordrecht: Kluwer.

Schmidt, R. A., Young, D. E., Swinnen, S. & Shapiro, D. C. (1989). Summary knowledge of results for skill acquisition: support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 352–359.

Schmidt, W. H., McKnight, C. C., Houang, R. T. *et al.* (2001). *Why Schools Matter: A Cross-National Comparison of Curriculum and Learning*. San Francisco, CA: Jossey-Bass.

Schofield, J. W. (1995). *Computers and Classroom Culture*. New York: Cambridge University Press.

Shavelson, R. J., Phillips, D. C., Towne, L. & Feuer, M. J. (2003). On the science of education design studies. *Educational Researcher*, *32*(1), 25–28.

Shavelson, R. J. & Towne, L. (2002). *Scientific Research in Education*. Washington, DC: National Academy Press.

Shea, J. B. & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 179–187.

Shonkoff, J. P. & Phillips, D. A. (eds.) (2000). *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington, DC: National Academy Press.

Simon, D. A. & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, *4*, 907–912.

Slamecka, N. J. & Graf, P. (1978). The generation effect: delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592–604.

Sloane, K. & Linn, M. C. (1988). Instructional conditions in Pascal programming classes. In R. E. Mayer (ed.), *Teaching and Learning Computer Programming: Multiple Research Perspectives*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Son, L. K. (2004). Spacing one's study: evidence for a metacognitive control strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 601–604.

Songer, N. B., Lee, H.-S. & McDonald, S. (2003). Research towards an expanded understanding of inquiry science beyond one idealized standard. *Science Education*, *87*(4), 490–516.

Stokes, D. E. (1997). *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington, DC: Brookings Institute.

Thios, S. J. & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior*, *15*, 529–536.

Thompson, C. P., Wegner, S. K. & Bartling, C. A. (1978). How recall facilitates subsequent recall: a reappraisal. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 210–221.

Tzeng, O. J. L. (1973). Stimulus meaningfulness, encoding variability, and the spacing effect. *Journal of Experimental Psychology*, *99*, 162–166.

Wheeler, M. A., Ewers, M. & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*, 571–580.

Wheeler, M. A. & Roediger, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, *3*, 240–245.

Whitten, W. B. & Bjork, R. A. (1977). Learning from tests: the effects of spacing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 465–478.

Winstein, C. J. & Schmidt, R. A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 677–691.