Introducing Desirable Difficulties for Educational Applications in Science (IDDEAS)

Robert A. Bjork *(University of California, Los Angeles)* and

Marcia C. Linn *(University of California, Berkeley),* Principal Investigators

APPLICATION ABSTRACT

Students' performance during instruction is commonly viewed as a measure of learning and a basis for evaluating and selecting instructional practices. Basic-research findings question that view: Conditions of practice that *appear* optimal during instruction can fail to support long-term retention and transfer of knowledge and, remarkably, conditions that introduce difficulties for the learner—and *appear* to slow the rate of the learning—can enhance long-term retention and transfer. Such "desirable difficulties" (Bjork, 1994, 1999) include spacing rather than massing study sessions; interleaving rather than blocking practice on separate topics; varying how to-be-learned material is presented; reducing feedback; and using tests as learning events.

The benefits of desirable difficulties found using simple laboratory tasks and short retention intervals not only raise concerns about prevailing educational practices, but also suggest unintuitive ways to enhance instruction. The present study focuses on whether such results generalize to realistic educational materials and contexts. In controlled experiments involving middle-school and college students, the effectiveness of standard Web-Based Inquiry Science Environment (WISE, http://wise.berkeley.edu) projects—on topics such as light propagation, thermal equilibrium, and science and treatment of malaria—will be contrasted with the effectiveness of experimental versions that incorporate selected desirable difficulties.

As a tool for teachers and students, WISE projects can enhance science education. If successful, this investigation can bridge the science of cognition and education and provide theoretically based principles that designers can use to create new materials.

Research Narrative

National Significance

Today, in this nation, we have both the will and the opportunity to upgrade education significantly. The will stems from a national consensus, and a constructive response by government agencies to that consensus. Political and business leaders, concerned parents, and typical citizens agree that America needs to improve education and enhance student achievement. Our educational system has significant shortcomings as shown in results from cross-national comparisons (e.g., TIMSS, 1998; Stigler & Heibert, 1999) and by studies of workplace competence (e. g., SCANS, 1991). Performance in mathematics and science and among students diagnosed as having disabilities has caused widespread alarm. There is also broad agreement that education *is* the future—not only for our children, but also for our nation as a whole.

The opportunity to upgrade education comes from extensive progress in the last several decades on understanding the cognitive processes that underlie learning. Basic research on learning and memory now provides a foundation for improving educational practices, potentially in revolutionary ways. Recent research, for example, questions the common view that student performance during instruction indexes learning and validly distinguishes among instructional practices. Work by Bjork and other researchers has established that conditions of practice that *appear* optimal during instruction can fail to support long-term retention and transfer of knowledge; whereas, and remarkably, conditions that introduce difficulties for the learner—slowing the *apparent* rate of the learning—can enhance long-term retention and transfer. (Such "desirable difficulties" (Bjork, 1994, 1999) include spacing rather than massing study sessions; interleaving rather than blocking practice on separate topics or tasks; varying how instructional materials are presented or illustrated; reducing feedback; and using tests rather than presentations as learning events.)

The opportunity to upgrade instruction also benefits from technology-enhanced learning environments that enable researchers to test the impact of these new research findings by consistently varying the conditions of instruction (e.g. Anderson, et al., 1996; 1997).

Enabled by the OERI Cognition and Student Learning program, we propose Introducing Desirable Difficulties for Educational Applications in Science (IDDEAS) to build bridges from the science of cognition to educational practices. IDDEAS will identify the laboratory-based principles and phenomena that do and do not generalize to educational settings and test mechanisms for implementing the principles and phenomena that *do* generalize in actual classrooms using technology-based instruction. IDDEAS requires a partnership of collaborating cognitive researchers, educational researchers, and classroom teachers who jointly design and carry out experiments in progressively more complex educational settings. We propose to form a sustainable partnership that can build a strong bridge linking the science of cognition, effective classroom practices, and powerful learning technologies. If successful, IDDEAS will develop theory-based principles to guide future instructional designers working in new contexts.

Project Design

The goal of our study is to examine the implications and potential of some unintuitive laboratory findings that seem to have particular promise for more complex academic learning. A key feature of our project design is the use of a set of existing web-based instructional modules as a test bed for the research we proposed. More specifically, the plan is to take advantage of the instructional modules that have been developed within the Web-based Inquiry Science Environment (WISE; http://wise.berkeley.edu), created and maintained by Marcia Linn and her collaborators at the University of California, Berkeley. WISE modules constitute a flexible and versatile educational tool for teachers and students, but they also have properties that make them an attractive research tool, as we outline in more detail below. Examples of the modules that we intend to use in the current research are those on the science and treatment of malaria, genetic modification of foods, light propagation, and thermal equilibrium.

The open nature of the WISE site makes it possible for researchers and teachers to augment and refine existing modules. Those same characteristics, together with the computer-based nature of WISE, makes those modules also an ideal test bed for the research we propose. Experimental and control versions of a given module can be contrasted without placing an added burden on teachers. The exportable nature of the modules means that they also have other important virtues as a research tool. They are reusable by other research groups, for example, and ideas can be tested using introductory-psychology students as well as middle-school students as participants. In addition, the results of this research can be easily shared, providing not only a basis for upgrading WISE modules and similar web-based instruction, but also for other forms of science education.

Motivating Considerations and Relevant Basic-Research Findings

In recent papers, Bjork and his collaborators (Bjork, 1994, 1999; Christina and Bjork, 1991; Ghodsian, Bjork, & Benjamin, 199X; Jacoby, Bjork, & Kelley, 1994; and Schmidt and Bjork, 1992) have argued that the typical instructional program is likely to be much less effective than it could be—because, basically, individuals responsible for the design of instruction are susceptible to being misled as to what are, and are not, effective conditions of learning. Conditions that enhance performance during the instructional process are assumed, implicitly or explicitly, to be the conditions of choice with respect to enhancing the goal of instruction:

namely, long-term post-instruction memory and comprehension. That assumption, however, in the light of a variety of laboratory findings, appears to be questionable at best and sometimes dramatically wrong. Manipulations that speed the rate of acquisition during training can fail to support long-term post-training performance, while other manipulations that appear to introduce difficulties for the learner during training can enhance post-training performance.

The Goals of Education

The most fundamental goals of education are long-term goals. As teachers and educators, we want targeted knowledge and skills to be acquired in a way that makes them durable and flexible. More specifically, we want a student's educational experience to produce a mental representation of the knowledge or skill in question that fosters long-term access to that knowledge and the ability to generalize—that is, to draw on that knowledge in situations that may differ on some dimensions from the exact educational context in which that knowledge was acquired. Verifying that someone has ready access to skills or knowledge in some standard situation does not, unfortunately, assure that that individual will be able to access that knowledge in a different situation, or on altered versions of the task in question. Even superficial changes can disrupt performance markedly. Perceived similarity, or the lack thereof, of new tasks to old tasks is a critical factor in the transfer of training (see, e.g., Gick & Holyoak, 1987).

Stated in terms of human memory, then, we would like a student's educational experience not only to produce a stored representation of some target knowledge in his or her long-term memory, but also to yield a representation that remains accessible (recallable) as time passes and contextual cues change.

Toward achieving the long-term goals of education, it is important to be conscious of some fundamental characteristics of humans as learners and rememberers. Humans do not, for example, store information in long-term memory by making any kind of literal recording of that information, but, rather, by relating that new information to what is already known--that is, to the information that already exists in memory. The process is fundamentally semantic in nature; information is stored in terms of its meaning, as defined by its associations and relationships to other information in our memories. The capacity for such storage is essentially unlimited--storing information, rather than using up memory capacity, appears to create opportunities for additional storage.

The process of accessing stored information given certain cues also does not correspond to the "playback" of a typical recording device. The retrieval of stored information is a fallible, probabilistic process that is inferential and reconstructive. Information that is readily accessible at one point in time, or in a given situation, may be impossible to recall at another point in time, or in another situation. The information in long-term memory that is, and is not, accessible at a given point in time is heavily dependent on the cues available to us, not only on cues that explicitly guide the search for the information in question, but also on environmental, interpersonal, mood-state, and body-state cues.

A final relevant important characteristic of human memory is that the act of retrieving information is itself a potent learning event. Rather than being left in the same state it was in prior to being recalled, the retrieved information becomes more recallable in the future than it would have been without having been accessed. In fact, as a learning event, it appears that a successful retrieval can be considerably more potent than an additional study opportunity, particularly in terms of facilitating long-term recall (see, e. g., Gates, 1917, Hogan & Kintsch, 1971, and Landauer & Bjork, 1978). There is also evidence that such positive effects of prior recall on the later recall of the retrieved information can be accompanied by impaired retrieval of competing information, that is, of other information associated to the same cue or set of cues as the retrieved information (see, e. g., Anderson, Bjork, and Bjork, 1994; Anderson & Spellman, 1995).

In a very general way, then, from the standpoint of our research-based understanding of human learning and memory, creating durable and flexible access to critical information in memory is partly a matter of achieving a certain type of encoding of that information, and partly a matter of practicing the retrieval process. On the encoding side, we would like the learner to achieve, for lack of a better word, an *understanding* of the knowledge in question, defined as an encoding that is part of a broader framework of interrelated concepts and ideas. Critical information needs to be multiply encoded, not bound to single sets of semantic or situational cues. On the retrieval side, practicing the actual production of the knowledge and procedures that are the target of training is essential. Similar to the argument for multiple encoding, it is also desirable to induce successful access to knowledge and procedures in a variety of situations that differ in the cues they do and do not provide.

The Need to Introduce (Desirable) Difficulties for the Learner

What specific manipulations of training, then, are best able to foster the long-term goals of training, whether stated in terms of measures of post-training abilities or in terms of underlying memory representations? Whatever the exact mixture of manipulations that might turn out to be optimal in a particular learning context, one general characteristic of that mixture seems clear: It would introduce more difficulties and challenges for the learner than are typically present in education. Surveys of the relevant research literatures (see, e.g., Christina & Bjork, 1991; Farr, 1987; Reder & Klatzky, 1994; and Schmidt & Bjork, 1992) leave no doubt that many of the most effective manipulations of training--in terms of post-training retention and transfer--

share the property that they introduce difficulties for the learner. Some of the clearest examples of such manipulations are the following.

(a) *Varying the conditions of practice*. Introducing variation and/or unpredictability in the training environment causes difficulty for the learner but enhances long-term performance--particularly the ability to transfer training to novel but related task environments. Where several differing motor-movement tasks are to be learned, for example, scheduling the practice trials on those tasks in random fashion, rather than blocking the trials by task type, has been shown to impair performance during training but enhance long-term performance (Shea & Morgan, 1979; Hall, Domingues, and Cavazos, 1992). Analogous results have been obtained with problem-solving tasks (e.g., Reder, Charney, & Morgan, 1986). Similarly, varying the parameters of a to-be-learned task--by, for example, varying the speed or distance of a target--impairs performance during training but enhances post-training performance (e. g., Catalano & Kleiner, 1984; Kerr & Booth, 1978). And the effects of increasing the variety, types, or range of exercises or problems (e.g., Carson & Wiegand, 1979; Gick & Holyoak, 1983; and Homa & Cultice, 1984) tend to exhibit the same general pattern. Even varying the incidental environmental context in which learning sessions are situated has been shown to enhance long-term retention (Smith, Glenberg, & Bjork, 1978; Smith & Rothkopf, 1984).

(b) *Providing contextual interference.* Such ways of making the task environment more variable or unpredictable can be considered one set of a broader category of manipulations that produce "contextual interference" (Battig, 1979). Other examples of contextual interference include designing or interleaving materials to be learned in a way that creates, at least temporarily, interference for the learner (e.g., Mannes & Kintsch, 1987), and adding to the task demands (e.g., Battig, 1956; Langley & Zelaznik, 1984). In Mannes and Kintsch's experiment, for example, subjects had to learn the content of a technical article (on industrial uses of microbes) after having first studied an outline that was either consistent with the organization of the article, or inconsistent with that organization (but provided the same information in either case). The inconsistent condition impaired subjects' verbatim recall and recognition of the article's content (compared to the consistent condition), but facilitated performance on tests that required subjects to infer answers or solve problems based on their general understanding of the article's content.

(c) *Distributing practice on a given task.* In general, compared to distributing practice sessions on a given task over time, massing practice or study sessions on to-be-learned procedures or information produces better short-term performance or recall of that procedure or information, but markedly inferior long-term performance or recall. The long-term advantages of distributing practice sessions over time has been demonstrated repeatedly for more than a century, tracing back over the entire history of controlled research on human memory (for modern reviews, see Dempster, 1990, 1996; Glenberg, 1992; and Lee & Genovese, 1988). In an experiment by Bahrick (1979), for example, the participants' basic task was to learn the Spanish translations of a list of 50 English words and the time between successive study/practice sessions 0, 1, or 30 days for different groups of participants. Looking at performance at the start of each study session, performance was clearly best with the 0-day separation, then the 1-day separation, then the 30-day separation, but on a final criterion test administered after 30 days for all groups, the pattern was dramatically reversed, with the 30-day spacing of training sessions yielding clearly superior recall.

(e) *Reducing feedback to the learner.* Until recently, a common generalization about motor skills was that providing external feedback to the learner facilitates the acquisition of skills, and that any means of improving such augmented feedback--by, for example, making it more immediate, more frequent, or more accurate--helps learning and performance. Recently, however, Richard Schmidt and his collaborators (see, e.g., Schmidt, 1991; Schmidt, Young, Swinnen, & Shapiro, 1989; Winstein & Schmidt, 1990) have found that--as in the case of the other manipulations summarized in this section--reducing the frequency of feedback makes life more difficult for the learner during training, but can enhance post-training performance. They have demonstrated that providing summary feedback to subjects (after every 5 or 15 trials, for example), or "fading" the frequency of feedback over trials, impedes acquisition of simple motor skills but enhances long-term retention of those skills.

(f) *Using tests as learning events.* Such effects of reducing the frequency of feedback during the learning of motor skills are broadly consistent with a large verbal-memory literature on tests as learning events. As mentioned earlier, there is abundant evidence that the act of retrieval induced by a recall test can be considerably more potent than a study opportunity in facilitating future recall. Prior testing also appears to increase the learning that takes place on subsequent study trials (e.g., Izawa, 1970). Once again, however, using tests rather than study trials as learning events, or increasing the difficulty of such tests, may appear to be counterproductive during training. Hogan and Kintsch (1971), for example, found that study trials produced better recall at the end of an experimental session than did test trials, but that test trials produced better recall after a 48-hour delay. And Landauer & Bjork (1978; see also Rea & Modigliani, 1985) found that "expanding retrieval practice," in which successive recall tests are made progressively more difficult by increasing the time and intervening events prior to each next test of some target information, facilitates long-term recall substantially--compared to the same number of tests administered at constant (and easier) delays.

What the foregoing desirable difficulties share in common is that they enable or require effective encoding and/or retrieval operations. They induce more elaborate encoding processes and more substantial and varied retrieval processes. As Battig (1979) argued with respect to contextual interference, and Schmidt and Bjork (1992) have argued more broadly, such manipulations are likely to induce more "transfer appropriate processing" (Bransford, Franks, Morris, & Stein, 1979; Morris, Bransford, & Franks, 1977)--that is, processing that will transfer to the post-training environment.

In summary, then, the research picture is that a variety of manipulations that impede performance during instruction facilitate performance on the long term and that pattern has the potential to mislead those responsible for the design of instruction. For evidence that the learner himself or herself is also susceptible to being fooled by his or her current performance, see Baddeley & Longman (1978) and Simon & Bjork (2001).

The Web-based Inquiry Science Environment (WISE)

Many researchers have observed that technology can support inquiry activities through software scaffolds, real-time display of data, online interactions, analytic tools, visualizations, simulations, and access to information through databases or Web sites (Edelson et al., 1999; Reiser et al., 1992; Scardamalia & Bereiter, 1992; Vanderbilt, 1993, 1997; Pea & Gomez, 1993; Linn & Songer, 1991; Linn, 1992; Songer, 1996). A common feature of this research has been the development of computer-based learning environments, such as Scientists in Action (Bransford et al., 1999) or the Knowledge Integration Environment (Linn, et al., 2000) and, as used in this research, the Web-based Inquiry Science Environment (WISE). Computer-based learning environments can make inquiry projects more successful by offering students cognitive and procedural guidance and freeing teachers to interact with students about complex science topics (Vanderbilt, 1997; Feurzeig & Roberts, 1999; Kafai & Resnick, 1996; Slotta & Linn, 1999; Linn & Hsi, 2000; White & Frederiksen, 1998). They can also enable embedded assessments that capture student understanding of science principles.

Educational researchers, district administrators, and even science standards call for integrating technology within the science curriculum, but far too many programs emphasize "basic skills training" for teachers (e.g., teaching them how to build a Web site or develop a PowerPoint presentation). WISE is designed to incorporate *fluency with information technology* ("FITness"), enabling students to collaborate with peers, test solutions using complex models, or search for current research findings (NRC, 1999). Surveys show that teachers agree with the importance of technology and inquiry but lack the resources and professional programs to incorporate these innovations into their teaching (Becker, 1999; Becker, Ravitz and Wong, 1999).

In the past three years, WISE has grown steadily, with dramatic recent gains. As the curriculum library has become larger and the learning environment technology more accessible, the project has reached a point where it is a viable approach for teachers who would like to incorporate inquiry and technology into their science curriculum.

Inquiry Instruction, Knowledge Integration, and Desirable Difficulties

Research on how individuals make sense of the natural world clarifies why inquiry instruction succeeds (Brown & Bransford, 2000; Piaget, 1971; Inhelder & Piaget, 1978?; Scardamalia & Bereiter, 1996; Vygotsky, 1962). Learners develop scientific expertise by interpreting the facts, processes, and inquiry skills they encounter in light of their own ideas and experiences. Typically, students hold a repertoire of ideas about scientific phenomena and investigations (Driver, 1985; Driver, Leach, Millar, & Scott, 1996; Eylon & Linn, 1994; Slotta, Chi, & Joram, 1995). Piaget (1971) highlighted the importance of the ideas that students bring to science class, and Vygotsky (1962) differentiated between students' instructed ideas and spontaneous ideas developed from personal experience. Inquiry instruction is successful because it engages these diverse ideas and challenges learners to engage in *knowledge integration:* a process where students make connections between their existing ideas, information introduced in science class, observations, and alternative perspectives suggested by peers or experimental investigations (Scardamalia & Bereiter, 1991; diSessa, 2000; Linn & Hsi, 2000).

Knowledge integration is a process that includes interpretive, cultural and deliberate aspects (Linn, in press). First, learners *interpret* new material in light of their own ideas and experiences, frequently relying on personal perspectives rather than instructed ideas. To take advantage of the interpretive nature of learning, inquiry instruction must enable students to organize, prioritize, and compare alternative perspectives, offering new and potentially pivotal cases that bring to light conundrums wherever possible.

There is a clear and compelling relationship between the ideas derived from prior work using the WISE environment and the learning and memory principles—derived from laboratory research by cognitive psychologists—summarized earlier in this proposal. The inquiry process challenges students in multiple ways—ways that introduce variability, require generation and retrieval practice, and tap into other components of effective processing. Achieving knowledge integration is akin to developing the kind of elaborated and inter-linked memory representation that basic researchers have shown will sustain access to knowledge, retard forgetting, and enhance transfer. The work by Linn and her collaborators on ”pivotal cases” (summarized in Linn, in press) is especially congruent with the desirable-difficulties framework. When such cases—which are observations and science facts that are difficult or impossible to for students to make sense of given a common misconception of one kind or another—are presented, students are confronted with a cognitive challenge, in effect. They must resolve, or attempt to resolve, a conflict between what a pivotal case implies versus what should be the case given their personal theory in a given domain.

WISE, Desirable Difficulties, and the Development of Autonomous Learning Skills.

The most fundamental goal of education is not the learning of content, per se, but fostering learning-to-learn skills and the motivation to continue learning across one's lifetime. An inquiry curriculum, such as that intrinsic to WISE, can enable the development of lifelong learning skills like critiquing evidence, debating arguments, or designing solutions to personally relevant problems. From a metacognitive standpoint, we want students to both acquire learning skills and to learn about the conditions that do and do not enhance long-term comprehension and transfer. Towards that goal, as argued above, they must learn that current performance is not a reliable guide to when comprehension and understanding has happened. It is necessary, in short, to become suspicious of the sense of ease and undeterred by the sense of difficulty.

Research plan

To bridge cognitive research and educational practice we will add desirable difficulties to WISE projects and conduct both experiments and quasi-experiments to test the conceptual framework. We call the quasi-experiments we plan to conduct compelling comparisons because they contrast two conditions of interest to policy makers and because they control for as many variables as possible in authentic school situations. We will use the results from these investigations to develop priniciples that enhance the conceptual framework for desirable difficulties. We will test these principles in a pilot study where we observe designers of complex educational materials when they use these priniciples and when they do not have these principles.

Adding desirable difficulties to WISE projects. We will use the conceptual framework for desirable difficulties to "enhance" a group of WISE projects in regular use in middle school and high school. Our partnership of psychologists, classroom teachers, and educators will identify promising ways to use desirable difficulties in WISE projects—which already address difficult science concepts and some might question the advantages of making these projects more challenging. We will focus on the most challenging aspects of these projects and identify potential ways to redesign these challenges in line with prior research on desirable difficulties. We will take care to assure that the redesigned projects require the same amount of classroom time as do standard projects. We will also keep the reading level of the materials and the homework demands identical in the two versions of the projects.

In what follows, we illustrate potential benefits of desirable difficulties for two WISE projects: *The Malaria Debate* and *How far does light go? The Malaria Debate* engages students in understanding the life cycle of the mosquito, the transmission of malaria, and the role of spraying with DDT, developing a vaccine, and instituting behavioral solutions, such as using bednets in controlling the disease. *How far does light go?* engages students in contrasting models of light propagation.

Our goal, across the life of this project, is to select most promising and readily adaptable desirable difficulties in our investigations. We illustrate the process for two desirable difficulties. First, we will study the advantages of interleaving content on different topics rather than presenting topics one at a time. For the *Malaria Debate*, for example, we will contrast the typical approach where students study the three alternatives in sequence with the enhanced approach where students encounter interleaved material about all three of the alternatives. We know that students will find the interleaved condition more "difficult" and we will test the impact of this difficulty on immediate and delayed posttests as well as on other outcome measures described below.

Second, we will study generation effects. For the *How far does light go?* project, we will create two conditions for students to investigate light intensity over distance and light reflection leading up to the research and argument construction on light propagation. In the typical (control) condition, students will read illustrated text on light propagation and reflection. In the enhanced (experimental) condition, students will be asked to generate predictions and design ways to investigate light intensity over distance and light reflection.

Selecting student participants. We will study the impact of typical and enhanced WISE projects in fully experimental designs using college students in the introductory-psychology research pool at the University of California, Los Angeles. Due to the constraints of class scheduling and student assignment to classes, we will conduct quasi experiments in middle schools and high schools (college students taking introductory psychology are typically only one or two years older than the high school students we intend to study). We will select pre-college courses that enroll a broad range of students, such as earth science and biology, to ensure that the students are representative of those in the school.

Selecting classroom teachers. We will collaborate with middle school and high school classroom teachers who have experience with WISE inquiry teaching, who teach two or more classes in the same subject, and who agree to use the same WISE unit in all the classes they teach on that subject. Teachers in these studies will be experienced, yet typical, of science instructors. Participating teachers will agree to random assignment of their classes to conditions, to use the project pretests and posttests for assessment purposes, and to administer the pretests and posttests the year prior to the year that they engage in the compelling comparison study in order to establish a baseline).

To study the impact of adding desirable difficulties to WISE projects we will select teachers who have previously taught the traditional WISE project at least twice. Our prior research (e.g. Slotta, in press; Linn & Slotta, in press) demonstrates that students in classes taught by teachers who have taught the WISE unit previously outperform students in classes taught by teachers using the WISE unit for the first time.

We will carry out compelling comparison studies where teachers use the traditional WISE project in half their classes and the WISE project enhanced with desirable difficulties in the other half of their classes. Our prior research demonstrates that classes differ one from another even when students are essentially assigned at random. Nonrandom assignment results from class scheduling, for example, as when the band meets during 5th period, meaning that these students cannot take science during that time. We will assign classes to conditions using a stratified random approach. We will ask teachers to group classes that appear similar in makeup and performance. We will randomly assign the grouped classes to conditions. In addition, to ensure comparability we will administer pretests and posttests. When necessary, we will use statistical approaches to adjust class outcomes on posttests based on pretest differences.

Assessing student learning. To assess student learning from WISE projects, we will carry out a number of activities. Students will participate in written pre-test and post-tests as part of instruction in each WISE project. In addition, each project includes embedded assessments that provide insight into student learning during instruction (See Figure 1). These form the primary outcome measures for the research.

To assess the generality of learning from WISE projects, we will focus on the success students have on their second WISE project, examining whether they learn the material more rapidly as well as more thoroughly as a result of experiencing an initial WISE project. To measure the generality of knowledge-integration instruction to more traditional measures, including state assessments as well as released items from state and national inventories, we will carry out two activities. First, we will identify released items from state and national tests relevant to WISE project and ask students in that school to use these items when they engage in test preparation for state examinations. We will request results from these items, which are relevant to the instruction students will receive, as part of our overall assessment plan.

Prior research on assessment in the context of WISE projects suggests that multiple choice and short answer questions are far less sensitive to WISE projects than are short and long essay assessments (Clark, 2000). As a result, we anticipate less impact on released items in state assessments than on WISE project pre-tests, post-tests, and embedded assessments.

To assess the validity of our findings, we will look at potentially confounding variables, such as prior science instruction and access to technology. We will use prior course experience, as well as state and school indicators, such as standardized test performance, and prior grades in science to identify students at risk for poor performance in science. We will study the relative advantages of desirable difficulties for students at risk and for those not at risk. In addition, we will focus specifically on access to technology as a potential factor in learning from WISE. Although we will ensure that students have adequate access to technology to complete WISE projects, we recognize that home access will vary and students with home or informal access outside of school may have advantages over those students without such access. To determine the potential advantage of this access, we will compare students with and without home access who are similar on other dimensions.

Year one research activities. During the first project year, we will design WISE projects with desirable difficulties and replicate psychological research designs using these more complex materials. We will study our own design process and identify what we conjecture to be preliminary principles based on our designs of enhanced WISE projects. We will compare this process to the pilot process described in year three.

Year-one college participants in WISE projects will come from the introduction-psychology research pool. We will invite students in groups of 4 to carry out WISE projects over a 6 hour [three session] period. We will replicate studies of interleaving material (malaria example given earlier) and studies of generation effects (light propagation example). [ADD DETAILS]. These studies are likely to suggest new questions that will be explored in year two and three, also using participants from the introductory-psychology pool.

For the classroom studies, we will collect baseline pretests and posttest from students in classes taught by the teachers who meet our criteria and agree to participate. Teachers will receive professional development and technical support from other funded research as well as from district sources.

Year-two research activities. During the second project year, we will revise the WISE projects based on results from college-student performance and identify the strengths and limitations of desirable difficulties based on the first year of investigation.

Year-two college student participants in WISE projects will study two new WISE projects enhanced with both new and previously studied desirable difficulties. Using the same desirable difficulty in two WISE projects is motivated by a need to determine whether instructional features do or do not generalize.

Pre-college participants during year two will come from the schools that carried out the baseline study in year one. During year two, teachers will have half their classes using enhanced WISE projects and half their students using typical WISE projects.

During year two we will also add schools in a primarily Hispanic district that collaborates with the WISE project. These students will participate in baseline pretests and posttests during year two. Teachers will receive professional development and technical support from other funded research as well as from district sources.

Year-three research activities. During the third project year we will revise the WISE projects introduced in year two based on results from college student investigations. We will synthesize the strengths and limitations of desirable difficulties based on two years of investigations and begin to frame principles that might be used by developers of new materials. We conduct a pilot project on the impact of these principles by asking a new cohort of WISE project developers to try these principles. We will conduct interviews with developers about their reactions. We can make informal comparisons between the activities of these developers and our own initial efforts to incorporate these principles into projects.

Pre-college participants during year three will come from all the schools that participated in year two. During year three, teachers in the first district will try using new WISE projects with desirable difficulties in half of their classes. That is, they will have half their classes using enhanced WISE projects and half their students using typical WISE projects.

During year three, pre-college participants will also come from the primarily Hispanic schools that carried out the baseline study in year two. Teachers will have half their classes using enhanced WISE projects and half their students using typical WISE projects.

Literature Cited

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, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. Psychological Review, 102, 68-100.

Baddeley, A.D., & Longman, D.J.A. (1978). The influence of length and frequency of training session on the rate of learning to type. Ergonomics, 21, 627-635.

Bahrick, H.P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. Journal of Experimental Psychology: General, 108, 296-308.

Battig, W.F. (1956). Facilitation and interference. In E.A. Bilodeau (Ed.), Acquisition of skill (pp. 215-244). New York: Academic Press.

Battig, W. F. (1979). The flexibility of human memory. In L. S. Cermak & F.I.M. Craik (Eds.), Levels of processing in human memory (pp. 23-44). Hillsdale, N. J.: Erlbaum.

Becker, Henry Jay. (February 1999). "Internet Use by Teachers: Conditions of Professional Use and Teacher-Directed Student Use." Center for Research on Information Technology and Organizations, The University of California, Irvine and University of Minnesota.

Becker, H., Ravitz, J. & Wong, Y. (1999). Teacher and Teacher-directed Student Use of Computers and Software. Report #3. Teaching Learning and Computing: 1998 National

Survey. Irvine, CA: University of California at Irvine, Center for Research in Information Technology in Organization.

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 peformance XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 435-459). Cambridge, MA: MIT Press.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). How people learn: Brain, Mind, Experience, and School. Washington, DC: National Research Council

Bransford, J.D., Franks, J.J., Morris, C.D., & Stein, B.S. (1979). Some general constraints on learning and memory research. In L.S. Cermack & F.I.M. Craik (Eds.), Levels of processing in human memory (pp. 331-354). Hillsdale, NJ: Erlbaum.

Bransford, J.D., Sherwood, R.D., Hasselbring, T.S., Kinzer, C.K. & Williams, S.M. (1990). Anchored instruction: Why we need it and how technology can help. In Nix, D. & Spiro, R. (Eds.), Cognition, education, and multimedia. Hillsdale: Erlbaum.

Carson, L. M., & Wiegand, R. L. (1979). Motor schema formation and retention in young children: A test of Schmidt's schema theory. Journal of Motor Behavior, 11, 247-251.

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.

Christina, R. W., & Bjork, R. A. (1991). Optimizing long-term retention and transfer. In D. Druckman & R. A. Bjork (Eds.), In the mind's eye: Enhancing human performance (pp. 23-56). Washington, DC: National Academy Press.

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

Dempster, F. N. (1990). The spacing effect: A case study in the failure to apply the results of psychological research. American Psychologist, 43, 627-634.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork and R. A. Bjork (Eds.), Memory: Vol. 10. Handbook of Perception and Cognition. (pp. 317-344). New York: Academic Press.

diSessa, A. (2000). Changing minds: Computers, learning, and literacy. Cambridge, MA, MIT Press.

Driver, R. (1985). Changing perspectives on science lessons. In N. Bennett, & C. Desforges (Eds.), Recent advances in classroom research. British Journal of Psychology Monograph.

Driver, R., Leach, J., Millar, R., & Scott, P. (1996). Young People's Images of Science. Buckingham, UK: Open University Press.

Edelson, D. C., Gordin, D. N., & Pea, R. D. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *Journal of the Learning Sciences, 8*(3/4), 391-450.

Eylon, B. S. and M. C. Linn (1988). Learning and instruction: An examination of four research perspectives in science education. Review of Educational Research 58(3): 251–301.

Farr, M. J. (1987). The Long-term Retention of Knowledge and Skills: A Cognitive and Instructional Perspective. New York: Springer-Verlag.

Feurzeig, W., & Roberts, N. (Eds.). (1999). Modeling and simulation in science and mathematics education. New York: Springer.

Gates, A. I. (1917). Archives of Psychology #40

Ghodsian, D., Bjork, R. A., & Benjamin, A. S. (1997). Evaluating training during training: Obstacles and Opportunities. In M. A. Quinones and A. Dutta (Eds.), Training for a rapidly changing workplace: Applications of Psychological Research (pp. 63-88). Washington DC: American Psychological Association.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. Cognitive Psychology, 12, 306-355.

Glenberg, A. M. (1992). Distributed practice effects. In L. R. Squire (Ed.), Encyclopedia of learning and memory (pp. 138-142). New York: Macmillan.

Hall, K. G., Domingues, E. & Cavazos, R. (1992). The effects of contextual interference on extra batting practice. Unpublished paper.

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.

Inhelder, B. & Piaget, J. (1958/1972). The growth of logical thinking from childhood to adolescence. New York: Basic Books.Izawa, 1970).

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

Jacoby, L. L., Bjork, R. A., & Kelley, C. M. (1994). Illusions of comprehensions and competence. In D. Druckman and R. A. Bjork (Eds.), Learning, remembering, believing: Enhancing human performance. (pp.57-80) Washington, DC: National Academy Press.

Kafai, Y. B., & Resnick, M. (Eds.). (1996). Constructionism in practice: Designing, thinking, and learning in a digital world. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kerr, R., & Booth, B. (1978). Specific and varied practice of a motor skill. Perceptual and Motor Skills, 46, 395-401.

Landauer, T. K., & Bjork, R. A. (1978). Optimal rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), Practical aspects of memory pp. 625-632). London: Academic Press.

Langley, D. J., & Zelaznik, H. N. (1984). The acquisition of time properties associated with a sequential motor skill. Journal of Motor Behavior, 16, 275-301.

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.

Linn, M. C. (1992). The Computer as Learning Partner: Can Computer Tools Teach Science? In K. Sheingold, L. G. Roberts, and S. M. Malcolm (Eds.), Technology for Teaching and Learning. Washington, DC: American Association for the Advancement of Science.

Linn, M. C. (in press). WISE design for lifelong learning—Pivotal Cases. In Gärdenfors, P. & Johansson, P. (Eds.), Cognition, Education and Communication Technology. Mahwah, New Jersey, Lawrence Erlbaum Associates.

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

Linn, M. C. (2000). "Designing the Knowledge Integration Environment: The partnership inquiry process." International Journal of Science Education, Special Issue (22): 781-796.

Linn, M.C. & Songer, N.B. (1991). Teaching thermodynamics to middle school students: What are appropriate cognitive demands? Journal of Research in Science Teaching, 28, 885-918.

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

Morris, C.D., Bransford, J.D., & Franks, J.J. (1977). Levels of processing versus transfer appropriate processing. Journal of Verbal Learning and Verbal Behavior, 16, 519-533.

National Research Countil (1999). High Stakes: Testing for Tracking, Promotion and Graduation. Washington, D.C.

Pea, R. D. & Gomez, L. (1993). Distributed multimedia learning environments: The Collaborative Visualization Project. Communications of the ACM, 36(5), 60-63.

Piaget, J. (1971). Structuralism. New York: Harper and Row.

Rea, C.P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. Human Learning, 4, 11-18.

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.

Reder, L. M., & Klatzky, R. L. (1994). Transfer: Training for performance. In D. Druckman and R. A. Bjork (Eds.), Learning, remembering, believing: Enhancing human performance (pp. 25-56). Washington, DC: National Academy Press.

Reiser, B.J., Kimberg, D.Y., Lovett, M.C., & Ranney, M. (1992). Knowledge representation and explanation in GIL, an intelligent tutor for programming. In J.H. Larkin & R.W. Chabay, Computer-assisted instruction and intelligent tutoring systems: Shared goals and complimentary approaches. Hillsdale: Erlbaum.

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(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., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. Psychological Science, 3, 207-217.

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.

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.

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 907-912.

Slotta, J.D., Chi, M.T.H. & Joram, E. (1995). Assessing the ontological nature of conceptual physics: A contrast of experts and novices. Cognition and Instruction.

Smith, S. M, Glenberg, A. M., & Bjork, R. A. (1978). Environmental context and human memory. Memory & Cognition, 6, 342-353. (

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. Cognition and Instruction, 1(3), 341-358.

Songer, N. B. (1996). Exploring learning opportunities in coordinated network-enhanced classrooms: A case of kids as global scientists. The Journal of the Learning Sciences, *5*, 297-327.

Slotta, J. & Linn, M.C. (2000). The Knowledge Integration Environment: Helping students use the Internet effectively. In M.J. Jacobson & R.Kozma (Eds.), Innovations in Science and Mathematics Education: Advanced Designs for Technologies of Learning (pp. 193-226). Mahwah: Erlbaum.

Vanderbilt, Cognition and Technology Group. (1997). The Jasper Project: Lessons in curriculum, instruction, assessment, and professional development: Lawrence Erlbaum Associates.

Vygotsky, L. S. (1962). Thought and language. Cambridge, MA, MIT Press.

White, B.Y. & Frederiksen, J.R. (1998). Inquiry, modeling, and metacognition: Making science accessible to students. Cognition and Instruction, 16, 3-118.

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.