Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings
Author(s): Ann L. Brown
Source: *The Journal of the Learning Sciences*, Vol. 2, No. 2 (1992), pp. 141-178
Published by: Taylor & Francis, Ltd.
Stable URL: http://www.jstor.org/stable/1466837
Accessed: 07/03/2011 03:59

THE JOURNAL OF THE LEARNING SCIENCES, *2*(2), 141-178

Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings

Ann L. Brown
University of California–Berkeley

The lion's share of my current research program is devoted to the study of learning in the blooming, buzzing confusion of inner-city classrooms. My high-level goal is to transform grade-school classrooms from work sites where students perform assigned tasks under the management of teachers into communities of learning (Bereiter & Scardamalia, 1989; Brown & Campione, 1990) and interpretation (Fish, 1980), where students are given significant opportunity to take charge of their own learning. In my current work, I conduct what Collins (in press) refers to as design experiments, modeled on the procedures of design sciences such as aeronautics and artificial intelligence. As a design scientist in my field, I attempt to engineer innovative educational environments and simultaneously conduct experimental studies of those innovations. This involves orchestrating all aspects of a period of daily life in classrooms, a research activity for which I was not trained. My training was that of a classic learning theorist prepared to work with "subjects" (rats, children, sophomores), in strictly controlled laboratory settings. The methods I have employed in my previous life are not readily transported to the research activities I oversee currently.

In Figure 1, I illustrate the critical aspects of my current classroom research. Central to the enterprise is that the classroom must function smoothly as a learning environment before we can study anything other than the myriad possible ways that things can go wrong. Classroom life is synergistic: Aspects of it that are often treated independently, such as teacher training, curriculum selection, testing, and so forth actually form

Requests for reprints should be sent to Ann L. Brown, University of California–Berkeley, Graduate School of Education, Education in Mathematics, Science, and Technology, Berkeley, CA 94720.

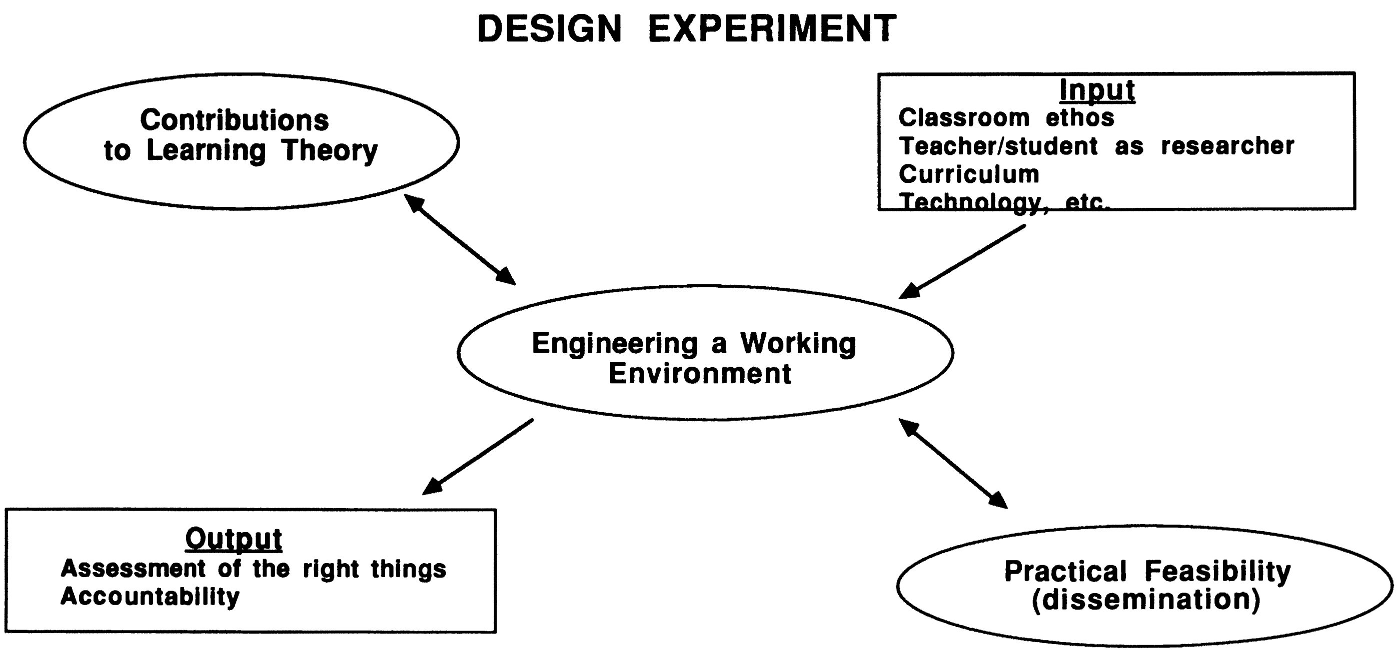

FIGURE 1 The complex features of design experiments.

part of a systemic whole. Just as it is impossible to change one aspect of the system without creating perturbations in others, so too it is difficult to study any one aspect independently from the whole operating system. Thus, we are responsible for simultaneous changes in the system, concerning the role of students and teachers, the type of curriculum, the place of technology, and so forth. These are all seen as inputs into the working whole.

Similarly, we are concerned with outputs from the system, a concern that leads us to look at new forms of assessment. It is essential that we assess the aspects that our learning environment was set up to foster, such as problem solving, critical thinking, and reflective learning. Assessment also allows us to be accountable for the results of our work to the children themselves, to parents, to teachers, to local authorities, and, last but not least, to fellow scientists.

Another critical tension in our goals is that between contributing to a theory of learning, a theoretical aim that has always been a keystone of our work, and contributing to practice. This is intervention research designed to inform practice. For this to be true, we must operate always under the constraint that an effective intervention should be able to migrate from our experimental classroom to average classrooms operated by and for average students and teachers, supported by realistic technological and personal support.

In the first part of the paper, I trace briefly how I moved from the classical psychological position of concentrating on a theoretical study of the learning processes of individual students to a concentration on conceptual change in teachers as well as students, to setting up a classroom ethos that would foster self-reflective learning, and finally to bigger concerns with technology, curriculum issues, and assessment.

My change in focus was a gradual evolution rather than an unpremeditated leap to instruction (Brown & Campione, 1986). Even though the research setting has changed dramatically, my goal remains the same: to work toward a theoretical model of learning and instruction rooted in a firm empirical base. I regard classroom work as just as basic as my laboratory endeavors, although the situated nature of the research lends itself most readily to practical application. In the classroom and in the laboratory, I attempt to engineer interventions that not only work by recognizable standards but are also based on theoretical descriptions that delineate why they work, and thus render them reliable and repeatable.

In the first part of this paper, I describe how I arrived at the current juncture, that is, attempting to restructure classrooms at many levels including designing curricula, introducing new roles for teachers, students, and researchers, and reconceptualizing assessment. The problem with such design experiments is that changes made in one part of the system reverberate throughout, and one is often forced to conduct the multiply confounded experiments that are a nightmare for an experimental psychologist. This discussion provides the groundwork for the major part of the article, in which I address the theoretical and methodological problems of working in the rich, complex, and constantly changing environment of the classroom.

LEARNING THEORY: A PERSONAL HISTORY

In this section, I trace my own shift toward studying learning in classrooms, although the steps I took reflect changes in the field in general. Beginning in the 1970s, and hence avoiding psychology's major shift away from behaviorist learning theories, I concentrate on the more subtle changes that took place in learning theory after the so-called cognitive revolution. Although the theoretical perturbations of this period were less traumatic than in the previous era, there was nothing short of a sea change in theories of learning, with a corresponding revolution in how learning was conceived, observed, and fostered. There was also a dramatic change in what "subjects" were required to learn, even in laboratory settings, and an awakening to the fact that real-life learning inevitably takes place in a social context, one such setting being the classroom. Psychologists are creatures of their time, and the methods they use to attack such durable problems as learning must be reconsidered in the light of theory change. The main point of this article is that learning theory has undergone major modifications even within the cognitivist period of the last 20 years, and methodological changes are needed to reflect these developments. I trace the sea change through a personal history.

Throughout my career, I have been interested in the problem of learning to learn, a topic with an honorable history in psychology and education. Educationalists have been interested in the problem because of two stumbling blocks to lasting learning: (a) inert knowledge (Whitehead, 1916): students acquire facts that they cannot access and use appropriately; and (b) passive learning: students do not readily engage in intentional, self-directed action. These two problems are commonly regarded as diseases of schooling (Brown, 1977). Attempts to study these phenomena led to a research agenda that saw fundamental changes in the focus of developmental theory, an agenda that focused on active strategies for learning, what it means to learn (rote learning vs. understanding), the content that learners are to acquire, and the context in which they are to acquire it.

Strategies, Capacity, and the Training Study

Like many experimental and developmental psychologists operating during the 1970s, my research focused on human memory: We avoided the term *learning* because of its association with behaviorist theories. During this period, there was an important shift in how we characterized children's purportedly poorer memories; we progressed from blaming poor performance on limited memory capacity per se to concentrating on the role of active memory strategies (Brown, 1974, 1975; Chi, 1976, 1978). This controversy led to the introduction of the *training study* as a theoretical tool to tease apart the relative contributions of capacity and strategic activity. The argument took the form of a question: Do young children fail to use strategies because they do not think to do so, or do they suffer from a mediational deficiency such that strategic intervention would not help them anyway? The response to this theoretical question was to train children in the use of strategies to see if their performance would be mediated effectively once they applied appropriate strategies. In information processing terms the distinction was between voluntary control processes (strategies) and structural features (memory capacity), a distinction that also led to attempts to distinguish the two through training (Atkinson & Shiffrin, 1968; Brown, 1975). Note, however, that this was an issue of theory, an attempt to unravel mysteries of the memory system. The work had nothing to do with theories about how to make a better learner. The players had absolutely no educational relevance in mind.

The main findings of 10 years' work can be summarized in two statements:

1. *Training worked,* even given the pathetic amounts provided in the typical study. Children could be trained to use simple strategies, and when they did use them, memory improved.
2. *Training didn't work,* because there was little evidence of maintenance in the absence of the experimenter's prompting and less evidence of independent transfer (Brown, 1978; Brown & Campione, 1978).

Important methodological issues sprang from this change of focus. When dealing with capacity, one simply asked how much was learned; this was true even when taking into account such complex problems as "chunking" (Chi, 1976; Miller, 1956). But cognitive strategies are typically conducted in the privacy of one's mind, and therefore it became necessary to devise creative means of externalizing these critical mental events. Extremely clever ways were devised to trick subjects into showing how they went about remembering, not just how much they retained. The focus on internal-memory processes rather than external memory products raised interesting measurement questions that remain problematic (Brown, Bransford, Ferrara, & Campione, 1983). And a recurring theme in this article is that theoretical changes during the last 20 years brought with them increasing demands for more subtle methods to render overt the human thought processes that are usually covert.

Metacognition

A major impetus to my early work was a review in which a leading theoretician characterized the vast body of memory research as trivial at best, and pointed out that psychologists do not study the most interesting thing about human memory, the fact that people have knowledge and beliefs about it (Tulving & Madigan, 1970). This topic became known as metamemory (Flavell, 1971), and it was enthusiastically studied by developmental psychologists because the problem of inert knowledge and passive learning appeared to be more prevalent in the young.

My research agenda in the 1970s was governed by these questions: Why is it that young children do not use strategies? Is it that they do not know that such activities are useful, or that they do not care, or a little of both? Enter metacognition: knowledge about and control of one's own learning. Studies of metacognition dominated the latter half of the 1970s, even though it was agreed that the term was fuzzy and ill defined (Brown, 1975, 1978; Flavell & Wellman, 1977; Wellman, 1983). Again, to summarize the outcomes crudely: Children do not use a whole variety of learning strategies because they do not know much about: (a) *remembering* (they know little about the strategies and tactics of overcoming memory limitations) and (b) *monitoring* (they do not think to orchestrate, oversee, plan, and revise their own learning activities). The terms remembering, monitoring, strategies, and metacognition widely replaced *memory,* denoting an important theoretical shift from passive to active metaphors of learning. This shift was attended by a need to develop new methods to capture active strategic and monitoring processes (Brown & Campione, 1979; Schoenfeld, 1985).

Again, along with others, I took the next step in this saga and attempted to train metacognition. The reasoning was as follows. If metacognition is the missing link, let's train it. To cut a long story short, it turns out not to be easy to train a learner to be strategic, to select cognitive activities intelligently, to plan, to monitor, to be cognitively vigilant, economical, and effective! And it is particularly difficult to do so in arbitrary contexts where a learner is attacking meaningless material for no purpose other than to please an experimenter. I believe it is for this reason that the decontextualized approach to metacognitive training was largely unsuccessful, and issues of the content and context of learning came to dominate the 1980s.

Content and contexts of learning. I will gloss over the history of these developments and merely point out that, concomitant with the emphasis on active learning, there were far-reaching modifications in what it was that children were required to learn, together with the contexts of learning in which they were observed. These developments also had a fundamental influence on the nature of learning theory.

During the 1970s, psychologists interested in learning and memory began shifting from the almost exclusive study of the learning of lists of words, pictures, and paired associates to a concentration on coherent content. They began also to look at the acquisition of expertise within a domain—gained over long periods of time via concentrated, and often self-motivated, learning (chess, for example). More recently, learning theorists have considered the acquisition of disciplined bodies of knowledge, characteristic of academic subject areas (mathematics, programming, physics, and even some areas of social science). Learning theories had to change to reflect this (Glaser & Bassok, 1989). My own research also reflected a shift to the study of learning, remembering, and understanding complex texts, which in turn led to studies of reading comprehension and comprehension monitoring in specific content areas.

These changes were accompanied by the need for novel methodologies. When studying list learning, one simply asked how many items were recalled, or, more subtly what organization is inherent in the order in which items are recalled. This organization was taken as a measure of conceptual level, preference, or proclivity. But when dealing with understanding of complex texts, the problem became one of how to capture degrees of understanding, nuances of meaning, acceptable alternative viewpoints, a problem that has kept the field busy for several years.

During the same period, there were also dramatic changes in the *contexts* in which students were asked to learn. For example, let us trace briefly what happened to the training study. The prototypical training study had the following features:

1. Training was absurdly cursory; few studies included more than a day or two of intervention. One might ask, what major cognitive change could be accomplished in a day or two?
2. The form of instruction was rarely discussed as a theoretical issue, and in general it was straight didactic teaching.
3. The interventions were typically one-on-one. Little thought was given to the social context of learning, or to collaborative cognition, which is so typical of everyday learning, including that of the classroom and the workplace.

To think of learning in this wider sense makes life more difficult, and again introduces nontrivial methodological problems. But changing views of learning in the 1980s have made that messy leap inevitable.

READING COMPREHENSION

These changes in learning theory led me down the slippery slope to educational theory and practice. Together with my colleagues, I began a series of studies concentrating on guided instruction and assessment in social contexts (Brown et al., 1983; Brown, Campione, Reeve, Ferrara, & Palincsar, 1991). In this article I concentrate on the most extensive of these programs, reciprocal teaching, a procedure designed to foster comprehension and cognitive monitoring while reading (Brown & Palincsar, 1982; Palincsar & Brown, 1984), solving mathematics problems (Campione, Brown, & Connell, 1988; Reeve, Gordon, Campione, & Brown, 1990), and learning science (Brown & Campione, 1990).

The primary work on reciprocal teaching centered on strategic reading. Although the focus was still theoretical—what was the role of strategies, metacognition, content, and context on learning and understanding—the applied value of the work was more apparent. The basic method was modeled after studies of Socratic or inquiry teaching, and theories about plausible reasoning, explanation, and analogy (Brown & Palincsar, 1989; Collins & Stevens, 1982). The procedure was designed to create a forum for externalizing simple comprehension-monitoring activities and to provide a repetitive structure to scaffold student discourse. (See Brown & Palincsar, 1989, and Palincsar & Brown, 1984, for details of the procedures.)

What happens in a typical reciprocal teaching session is that the participants in a group take turns leading a discussion about a text. The participants divide themselves up into a learning leader (adult or child) and learning listeners/critics for each segment of text. All group members get a turn as learning leader, who begins by *asking a question* and ends by *summarizing* the gist of what has been read. The group rereads and discusses possible problems of interpretation when necessary. Questioning provides the impetus to get the discussion going. Summarizing at the end of a period of discussion helps students establish where they are in preparation for tackling a new segment of text. Attempts to *clarify* any comprehension problems that might occur arise opportunistically, and the leaders ask for *predictions* about future content. These four activities—questioning, clarifying, summarizing, and predicting—were selected to bolster the discussion because they are excellent comprehension-monitoring devices; for example, if one cannot summarize what one has read, it is an indication that understanding is not proceeding smoothly and that remedial action is called for. The strategies also provide the repeatable structure necessary to get a discussion going, a structure that can be faded out when students are experienced in the discourse mode.

Over time our studies of reciprocal teaching of reading comprehension became more complex as the method became increasingly incorporated into the dynamics of a functioning classroom. We began working one-on-one with children who were reading unconnected passages in laboratory settings (Brown & Palincsar, 1982) and progressed to studying children in groups in resource rooms outside the classroom (Palincsar & Brown, 1984), then to considering naturally occurring reading groups in the classroom (Brown & Palincsar, 1989), and finally to studying reading comprehension groups that were fully integrated into science classrooms (Brown & Campione, 1990). We began by concentrating on a few constrained strategies and proceeded to study complex explanation, argument, and discussion forms. Initially we looked at students reading unconnected passages, proceeded to look at students reading coherent content (Brown et al., in press; Palincsar, Brown, & Campione, 1990), and we are now observing reading comprehension as it takes place in social groups who are reading, discussing, and arguing about cohesive material that they have prepared themselves over long periods of time, thus acquiring ownership of that knowledge. Finally, the students themselves appropriated reciprocal teaching as a tool to enhance their comprehension. Faced with important material they had difficulty understanding, they would call for a reciprocal teaching session to enhance and monitor their own comprehension. Similarly, before teaching their material to others, they would conduct reciprocal teaching sessions on it to ensure that it was comprehensible. Thus, in our current classrooms, reciprocal teaching sessions are initiated by children opportunistically when needed to ensure comprehension.

COMMUNITIES OF LEARNERS

In our current work, reciprocal teaching is only one component of the design experiment intended to encourage distributed expertise in a community of learners (Brown et al., in press). The settings are intact science classrooms where sixth-, seventh-, and eighth-grade students are responsible for doing collaborative research and sharing their expertise with their colleagues.

In order to create a community of learners, we must set up a classroom ethos that differs from that found in traditional classrooms. The main differences are indicated in Table 1. Although presented as dichotomous, the distinctions should be regarded as ends of continua. As bald dichotomies, they represent stereotypes.

In the traditional classroom, students are seen as relatively passive receivers of wisdom dispensed from teachers, textbooks, or other media. In the intentional learning classroom (Brown & Campione, 1990; Scardamalia & Bereiter, 1991), students are encouraged to engage in self-reflective learning and critical inquiry. They act as researchers responsible to some

TABLE 1
Changes in Classroom Philosophy

Role	*Traditional Classroom*	*Intentional Learning Environment*
Students	Passive recipients of incoming information	Students as researchers, teachers, and monitors of progress
Teachers	Didactic teaching	Guided discovery
	Classroom manager	Model of active inquiry
Content	Basic literacy curriculum	Thinking as basic literacy
	Lower vs. higher skills	
	Content curriculum	Content curriculum
	Breadth	Depth
	Fragmented	Recurrent themes
	Fragmented	Explanatory coherence
	Fact retention	Understanding
Computers	Drill and practice	Tools for intentional reflection
	Programming	Learning and collaboration
Assessment	Fact retention	Knowledge discovery and utilization
	Traditional tests	Performance
		Projects
		Portfolio

extent for defining their own expertise. Teachers' roles also change dramatically in that they are expected to serve as active role models of learning and as responsive guides to students' discovery processes. They teach on a need-to-know basis responsive to students' needs, rather than on a fixed scope and sequence schedule, or according to an inflexible lesson plan (Brown & Campione, 1990, in press). The content of the curriculum through which children are guided features a few recurring themes rather than breadth of coverage, themes that students come to understand deeply and recognize at increasingly deeper levels of explanatory coherence and theoretical generality. The technological environment is not designed to foster drill and practice or even programming, but to encourage intentional learning, reflection and communication. Finally, methods of assessment focus on the students' ability to discover and use knowledge, rather than just retain it. On-line dynamic measures of performance are as important as static measures of product (for more details of the actual classroom activities and outcomes, see Brown & Campione, 1990, in press; Brown et al., in press).

In order to foster a community of learners that features students as designers of their own learning, we encourage students to be partially responsible for creating their own curriculum. The two major forms of cooperative learning used to accomplish this are the Jigsaw method (Aronson, 1978) and reciprocal teaching. Students are assigned curriculum themes (e.g., animal defense mechanisms, endangered species, changing populations, food chains, etc.) each divided into five subtopics (e.g., changing populations = extinct, endangered, artificial, assisted, and urbanized populations; or food chains = producing, consuming, recycling, distributing, and energy exchange). Students form five *research groups,* each assigned responsibility for one of the five subtopics. These research groups prepare teaching materials using state-of-the-art computer technology (Campione, Brown, & Jay, in press). Then, using the Jigsaw method, the students regroup into *learning groups* in which each student is expert in one subtopic, holding ⅕ of the information. Each one-fifth needs to be combined with the remaining fifths to make a whole unit, hence "jigsaw." The expert on each subtopic is responsible for guiding reciprocal teaching seminars in his or her area. Thus, the choice of a discussion leader is now based on expertise rather than random selection, as was the case in the original reciprocal teaching work. All children in a learning group are expert on one part of the material, teach it to others, and prepare questions for the test that all will take on the complete unit. During this cyclical process, students both acquire content expertise and learn how to learn from texts and other media (Brown & Campione, 1990).

The students are involved in (a) extensive reading in order to research their topic, (b) writing and revision to produce booklets from which to teach and to publish in-class books covering the entire topic, and (c) computer use to publish, illustrate, and edit their booklets. In addition, a great deal of cognitive monitoring must take place in order for the students to set priorities concerning what to include in their books, what to teach, what to test, how to explain mechanisms, and so forth. They read, write, discuss, revise, set priorities, and use computers all in the service of learning. And they generate data!

I do not have space to describe the multiple types of data we gather. It is sufficient to say that in addition to relatively standard outcome measures involving reading, writing, content knowledge, and computer competence, all of which improve significantly (Brown & Campione, in press), this project generates a vast amount of information that is not readily subjected to standard measurement devices. We collect transcripts of children's planning, revising, and teaching sessions. We collect observations of teachers' coaching and responsive teaching, as well as their direct instruction. We have records of student portfolios, including individual and group long-term projects that require them to exploit accumulating knowledge in novel ways. We score electronic mail queries to peers, teachers, and collaborators in the university community. Ethnographic observations of cooperative, or not so cooperative, interactions, such as group discussions, planning sessions, help seeking, peer tutoring and so forth are taken routinely, together with extensive video and audio taping of individuals, groups, and whole classroom settings. In fact, we have no room to store all the data, let alone time to score it.

In the remainder of this article I concentrate on the methodological issues that arise when one attempts to situate a research agenda in a classroom setting that one is simultaneously involved in designing under conditions of continuous flux.

METHODOLOGICAL ISSUES IN EVALUATING COMPLEX INTERVENTION STUDIES

In this section, I consider some of the persistent problems associated with attempts to assess conceptual change in situ. I will concentrate on three main methodological issues: (a) The relationship between laboratory and classroom work, (b) Idiographic versus nomothetic approaches, or the grain-size issue, and (c) The Bartlett Effect, or the problem of data selection.

Classroom Versus Laboratory Contexts

The first issue I address is the tension between and relative contributions of classroom and laboratory studies. For most of my professional career, I have studied children learning in laboratory settings, usually in isolation from other children. Gradually over the years I have increasingly situated my study of learning in classrooms, first in such lab-like settings as pull-out time (for reading groups, etc.), then in socially sanctioned settings in the classroom (reading group), and finally orchestrating, some might say disrupting, the entire classroom activity for at least one hour a day. Making this shift involves an increasing trade-off between experimental control and richness and reality. The classroom is not the natural habitat of many experimental psychologists, and our methods did not evolve to capture learning in situ. Indeed, the first grant proposal I ever had rejected was about 10 years ago, when anonymous reviewers accused me of abandoning my experimental training and conducting "Pseudo-experimental research in quasi-naturalistic settings!" This was not a flattering description of what I took to be microgenetic/observational studies of learning in the classroom.

As a personal research strategy, I find that switching back and forth from both types of research settings enriches my understanding of a particular phenomenon. I give but one example. I have studied children's analogical reasoning and explanation strategies under tight experimental control (Brown & Kane, 1988) and as it occurs spontaneously in classroom discussions (Brown et al., in press). My observations go in both directions; my laboratory work informs my classroom observations—and vice versa.

My laboratory experiences enable me to see a developmental pattern emerging in classroom dialogues similar to that found in laboratory problem solving; for example, there is a progression from merely noticing the occurrence of analogies to productive use of analogy to solve problems (Brown & Kane, 1988; Gick & Holyoak, 1980). Similarly, there is a familiar progression from using surface similarities to a reliance on deep analogies (Gentner, 1983, 1989). For example, children initially notice the analogy between two passages where the protagonists (ladybug and lacewing) look alike, are both insects, and are examples of a common theme, for example, natural pest control. Next they notice the deep similarity between a crested rat and a viscount butterfly (very dissimilar animals who are examples of a common theme, visual mimicry). Similarly, children initially think the best analogy between a car and a human body is the eyes and the headlights (surface) whereas later they think that a better analogy is between the engine and the heart (deep). With increasing knowledge children progress from accepting superficial analogy to using deep analogy to explain mechanisms. They question initially acceptable surface analogies (such as "plant stems are like straws") and come to prefer analogies based on deeper understanding of underlying biological mechanisms ("plants are food factories"). This progression, reflecting the increasingly coherent and mechanistic nature of their biological theories, is shown in Figure 2. Although in the laboratory this development from noticing to using, and from surface to deep, was thought to be age-dependent, the classroom work suggests that the shift is knowledge-based, occurring microgenetically within a year as readily as cross-sectionally across several years.

Classroom work can motivate laboratory practice also: Trends discovered in spontaneous classroom discussions can be tested in the laboratory under more controlled conditions. For example, analysis of classroom dialogues suggests that the conditions of spontaneous use of explanation may be developmentally sensitive. First, impasse-driven explanation occurs in the face of breakdowns in comprehension. This is followed by the use of explanation to help resolve annoying inconsistencies. Finally, spontaneous explanation is used in the absence of comprehension failure or obvious inconsistencies as learners continually revise and deepen their understanding of complex causal mechanisms. This microgenetic progression is shown in Figure 3.

Faced with this apparent trend in the spontaneous dialogues, our routine procedure is to set up controlled laboratory studies to evaluate whether the developmental trend can be reproduced under experimental control. Similarly, faced with an experimental finding in the laboratory, we are sensitized to watch for its occurrence in the morass of classroom discourse. This

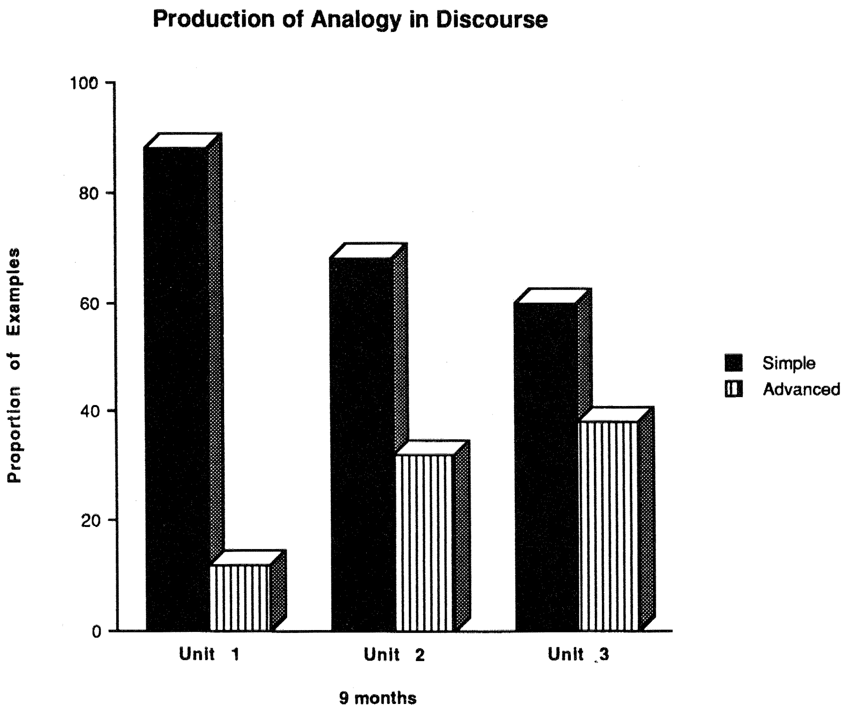

FIGURE 2 Production of analogy in spontaneous discourse of 5th and 6th graders during reciprocal teaching.

cross-fertilization between settings enriches our understanding of the developmental phenomenon in question. I would like to point out, however, that I regard neither aspect of the work as basic or applied. Theoretical advances can emerge from both the laboratory and classroom settings. They are just that, different settings whose features must be included in the description of the data they produce.

Idiographic Versus Nomothetic Approaches

One of the major decisions psychologists must make is to decide on the "grain size" of the phenomena they wish to observe. The classic distinctions are laid out neatly in introductory textbooks. Developmental psychologists are informed of the strengths and weaknesses of several such options. First there is the choice between an idiographic and a nomothetic approach, that is, few or many. Does one want to study a "single variable in many subjects for the purpose of discovering general laws or principles of behavior" or

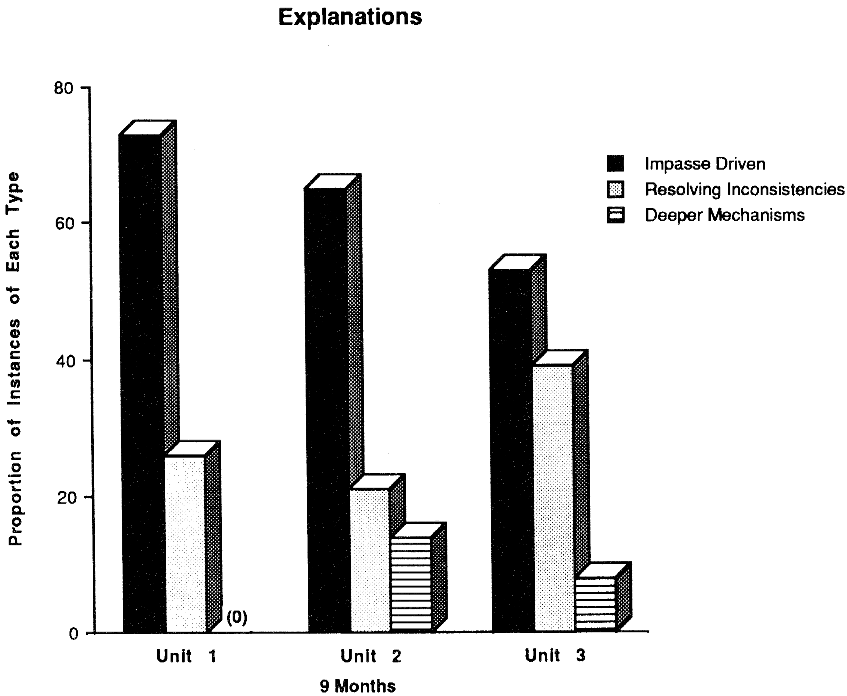

FIGURE 3 Production of explanation in the spontaneous discourse of 5th and 6th graders during reciprocal teaching.

adopt the idiographic approach, "the thorough study of individual cases, with emphasis on each subject's characteristic traits."

Although nomothetic studies are more typical in developmental psychology, case studies have an honorable history too, particularly in the study of very young children. The most famous of these studies is Darwin's (1877) delightful daily record of his eldest son. "So-called" diary studies recording infant development were popular at the turn of the century (Kessen, 1965); primarily conducted by mothers, they generated a great deal of the early language development data. More famous was the work of fathers, with Piaget (1952) studying his three children, Laurent, Lucienne, and Jacqueline, following the lead of Binet's (Binet, 1903) observations of his daughters, Madeleine and Alice. Large scale nomothetic studies were largely an American invention.

The second major variable has to do with change over time, a critical element for all learning theorists and especially for developmental psychologists. One major option is between cross-sectional designs, where data are taken from, say, 7-, 10-, and 14-year-olds, and inferences about developmental trajectories made; and longitudinal designs, where, for example, all 4-year-olds who watched "Sesame Street" in 1969, 1971, and 1973 are followed up for a period of years. Another option is the microgenetic design, an important tool in the psychologist's kit that is receiving renewed interest. Here children are observed over a relatively short period of time (days, weeks) as they acquire a certain form of understanding. This approach is most often taken with very young children who are at a stage of rapid learning (DeLoache, Brown, & Kane, 1985) but also with older children learning a particular scientific concept (Karmiloff-Smith & Inhelder, 1974–1975; Kuhn, Amsel, & O'Loughlin, 1988).

The third major decision is the type of analysis conducted, qualitative or quantitative. This decision is often partially dictated by the prior choices: Complex quantitative analyses are more often applied to large scale longitudinal and cross-sectional data bases but not always. Although theorists as clever as Binet, Piaget, and Skinner never used statistics, it is perfectly possible to subject case studies to statistical analysis if one should choose to do so.

Increasingly, I find that in the interest of converging operations, and because of the multifaceted nature of my data base, I prefer a mixed approach, suiting the method to the particular data. I mix and match qualitative and quantitative methodologies in order to describe the phenomena, a mixture that is becoming commonplace in the journals, reflecting the increasingly complex issues that psychologists now address. In my own work, I routinely combine a concentration on large scale data bases with in-depth microgenetic analyses of a few children or perhaps a group (Brown et al., in press; Campione, Rutherford, Gordon, & Brown, in press). At any one stage my laboratory is engaged in studying three to four classrooms in depth for a period of at least a year (a 3-year longitudinal study is currently underway). This means that typically we are working with approximately 100 students per year. For each student, we generate an enormous amount of data. Our routine practice is to take fairly traditional pretest and posttest data from all the experimental and control students and combine that with in-depth analyses of some of the students, complemented with a few selected case studies. For example, in the initial study of reciprocal teaching (Palincsar & Brown, 1984), we provided pretest and posttest data on the 37 participating students, mini case studies on six children, together with transcripts from two children who differed considerably in how quickly they picked up the reciprocal teaching procedure, a precedent we have followed in more recent work (Brown & Palincsar, 1987, 1989). This approach enables us to see the magnitude of the effect in terms of outcome measures and to get a feel for the phenomenon itself by looking at a particular child or group in depth. (Brown et al., in press; Campione et al., in press).

To give just one example of this mixed approach, consider our treatment of content knowledge acquisition. We give short answer quizzes on a pretest and posttest basis on all units of covered content. An example of such data is shown in Figure 4. Here the performance of approximately 60 fifth- and sixth-grade children in the experimental research classrooms is compared with a partial control group (PCont) that was treated exactly the same as the research classroom for the first semester (Unit 1), and then taught environmental science by their regular science teacher. The partial control group had exactly the same access to books, videos, computers, and so forth as did the research classrooms. As can be seen, the two groups did not differ from each other in Unit 1, where they were, in effect, both experimental groups, treated exactly the same; but the children in the research classrooms outperformed the partial control group on both Units 2 and 3. A read only control (ROC) group, who read the key materials but did no research, did poorly throughout. This traditional use of static pretests and posttests, combined with appropriate control data, provides us with clear evidence of the effectiveness of the intervention and is easy to share with school personnel as well as fellow scientists.

A richer picture of knowledge acquisition, however, comes from more in-depth probing of a subset of students. For example, we consider students' responses to clinical interviews and transfer tests. Students differ in their level of understanding and in the confidence with which they hold opinions; in the clinical interview it is the skilled interviewer's job to pose questions that are sensitive to the bandwidth of competence within which each individual student can navigate (Brown & Reeve, 1987). This is a form of dynamic assessment (Campione & Brown, 1987, 1990) that we use to measure emergent competence in general (see Campione et al., in press, for dynamic assessment of computer use in the community of learners). The point is to map the zone of proximal development (Vygotsky, 1978), the region of competence that a student can traverse with and without aid.

To map this window of opportunity for learning, the interviewer raises a series of key questions concerning, for example, the food chain, adaptation, and so forth. First, the interviewer elicits basic expository information (What does the student know about photosynthesis?). If the student cannot answer, the interviewer provides hints and examples as necessary to test the student's readiness to learn that concept. If, however, the student seems initially knowledgeable, the experimenter might question that understanding by introducing counterexamples to the student's beliefs (Is a mushroom a plant? What about yeast?), and, again if appropriate, he or she might ask the student to engage in thought experiments that demand novel uses of the information. For example, when a student has sorted pictures perfectly into herbivores and carnivores, and provided a good description of the difference, he or she may be asked, "What would happen on the African plain if

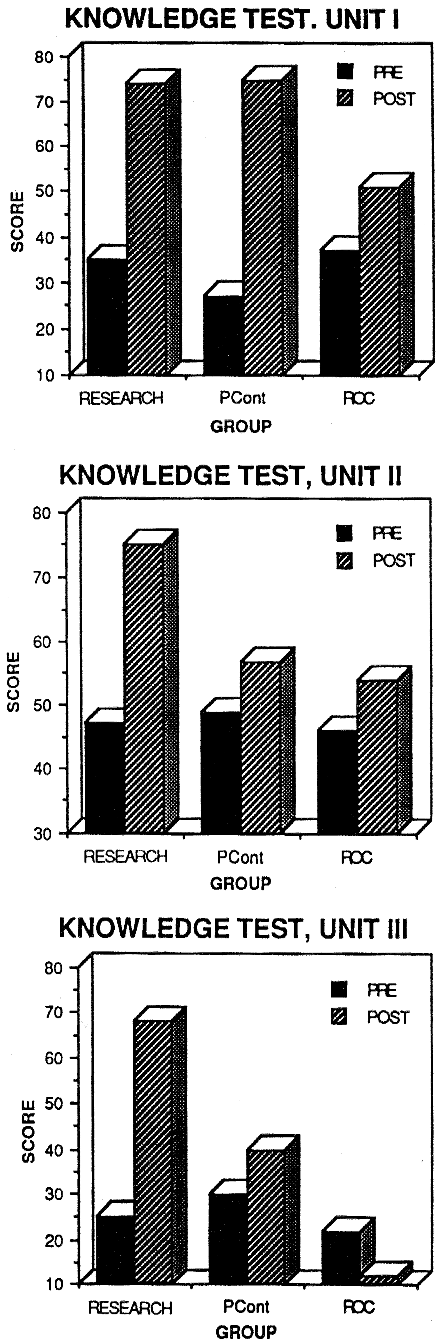

FIGURE 4 Pre- and posttest short answer knowledge tests for three curriculum units of environmental science. The data are from "Restructuring Grade School Learning Environments to Promote Scientific Literacy" in *Restructuring Learning: Analysis and Recommendations of the Council of Chief State School Officers* by A. L. Brown and J. C. Campione, in press, Washington, DC: Council of Chief State School Officers. Copyright by Ann L. Brown. Adapted by permission.

there were no gazelles or other meat for cheetahs to eat? Could they eat grain?" Some students are surprisingly uncertain about this, and may suggest that cheetahs could eat grain under certain circumstances, although they would not live happily. Some even entertain a critical period hypothesis—that the cheetah could change if it were forced to eat grain from infancy, but once it reached adolescence, it would be too set in its ways to change. Only a few invoke notions of form and function, such as properties of the digestive tract, to support the assertion that cheetahs could not change within their life span. These extension activities of thought experiments and counterexamples are far more revealing of the current state of students' knowledge than their first unchallenged answers, which often provide an overly optimistic picture of their knowledge.

For example, Katy, a sophisticated seventh grader, gave a textbook-perfect description of photosynthesis that would certainly be taken as an indication that she fully understood the basic mechanisms. She was then asked, "What would happen if there were 30% less sunlight?" Katy's response never included the critical information that as plants make food with the sun's energy, serious reduction in the availability of sunlight would disrupt the entire food chain—no sun, no plants, no food! Instead she concentrated on light to see with:

> That would kill off the plants, beetles, and, um, nocturnal things would be OK. The dayturnal things—snakes, rabbits, hares—would be alright, could be nocturnal. But the dayturnal things would need sunlight to see—couldn't find their food in the dark and would eventually starve to death. Hawks would also die out, but owls are nocturnal—would be able to see at night and, um, raccoons would probably be near the top of the food chain.

Katy clearly had not understood the basic place of photosynthesis as the mainspring of life. She can repeat back the mechanisms and form food chains when asked directly, but she cannot yet reason flexibly with her newfound knowledge.

Using these thought experiments, we can track not only retention of knowledge, but also how fragile-robust it is and how flexibly it can be applied. Consider the following excerpts from John, a sixth grader. During the pretest interview, John mentioned speed, body size, mouth size, and tearing teeth as functional physical characteristics of carnivores. He seemed to have the carnivore/herbivore distinction down pat. But when asked the cheetah thought experiment (Could they survive if there were no game?), he mused, ". . . well I mean if people can, like, are vegetarians, I mean I think a cheetah could change. . . ." When asked how this might happen, he said,

> Well . . . just to switch off, but, um, it would be easier for them to change on to plants than for me; if I had been eating meat . . . because there would still

> be meat around for me to eat, but for them there wouldn't be . . . so if they wanted to survive, they're going to have to eat grass.

When asked if it would be easier for a baby cheetah to eat grass, his response was:

> Well, if it was a baby, it would be easier because it could eat it . . . it would be right there, it would just have to walk a little bit to get it . . . but I think it would be easier . . . but then if it does happen for a long time, then the animals come back [the gazelles return], then the cheetah probably would have lost its speed, because they wouldn't have to run. . . . Yeah, and they'd get used to the grass and not care about the gazelle, because along the line they would forget.

During a subsequent interview 6 months later, when asked about herbivores and carnivores, John made complex analogies to the cow's intestinal system, and argued that herbivore digestive tracts are more complicated than those of carnivores. By knowing an animal's diet, he argued, he would be able to predict its digestive tract length and how long digestion might take, and vice versa.

This time, when confronted with a variant of the cheetah thought experiment, John responded, "No . . . no, their digestive system isn't good enough . . . it's too uncomplicated to digest grasses and also their teeth wouldn't be able to chew, so then the grass would overpopulate . . . and the cheetah dies. . . ." When asked if the baby cheetah could survive by eating grass, John asserted that they would be the first to die.

These responses are in distinct contrast to those given to the same questions during the pretest. John dropped personification (Hatano & Inagaki, 1987) as a justification ("humans can do it so cheetahs can too") and replaced it with a form-function response. Thrown a twist on the old question—would deer be able to eat meat if there were no longer grass—the newly confident John favored the interviewer with a broad smile and said, "Nice try—the digestive tract of the deer is too complicated and also the teeth wouldn't be able to grind meat."

Students in the research classroom show considerable improvement on such clinical interviews over the course of a year's study, in contrast to students taking a more traditional course that touches on the same materials (the partial control group).

We see the same benefit of being in the research classroom on application questions, such as "design an animal to fit the following habitat" (desert, tundra, or rain forest habitat described). Data from two replications of this test (at 6 months and 1 year) are shown in Figure 5. The sixth graders in the experimental classroom outperformed control students in the number of biologically-appropriate mechanisms that they included in their animals (such as mechanisms of defense, mechanisms of reproductive strategies, etc.). Over time the experimental subjects introduced more novel (untaught

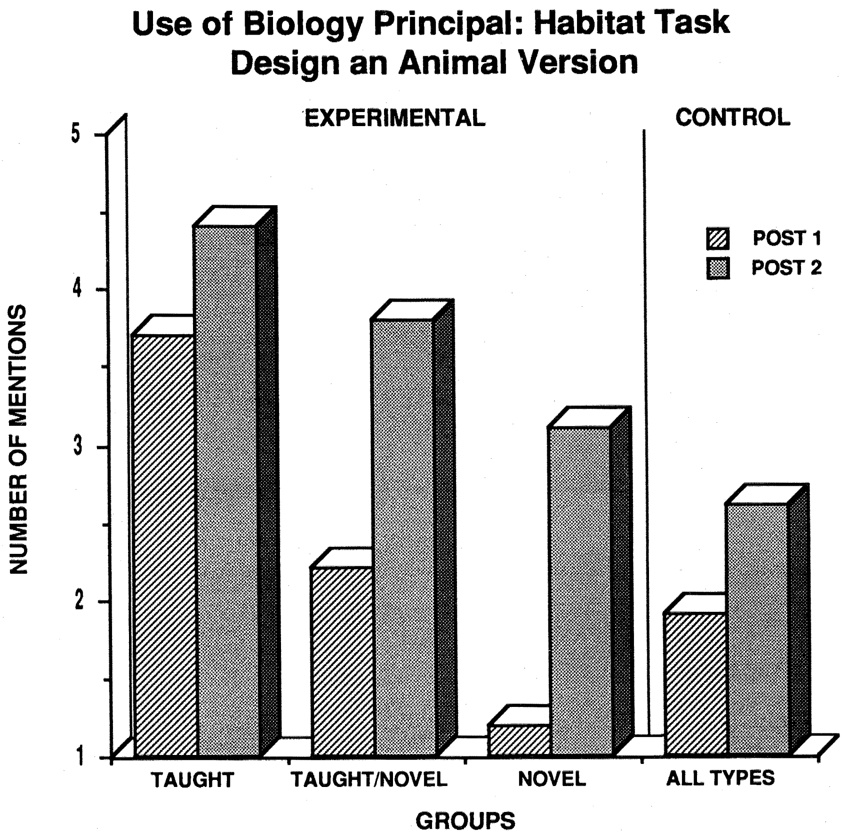

FIGURE 5 Use of taught versus novel biological principles in the design an animal transfer test.

mechanisms) and more novel variations of mechanisms they had been taught. For example, the class had discussed the notion of mimicry as a defense mechanism. In a response scored taught/novel, one student said that the eggs of his animal were placed in a line and the markings made the eggs look like a full grown cobra, a novel use of the mimicry principle which to our knowledge does not exist in nature.

In a variety of such clinical interviews and tests of application and transfer, the research children outperformed control groups in their ability to apply biological principles to novel tasks, such as designing a habitat to support an animal, or predicting systemic changes given one disruption in a food chain or ecosystem, activities that tap their accumulating knowledge in novel ways (Brown, et al., in press). Responses to clinical interviews and application activities are subjected to qualitative as well as quantitative analyses and tell us a great deal about the status of the child's accumulating knowledge and ability to reason on the basis of incomplete knowledge.

The Bartlett Effect

The third methodological issue, which goes beyond that of qualitative versus quantitative analyses, is that of data selection, particularly acute when portions of edited transcripts or clinical interviews are selected to illustrate a theoretical point, or when descriptions of planning sessions, peer tutoring, or teacher coaching are culled from a vast array of potential examples. With access to daily ethnographic notes, teacher's logs, and video and audio tapes, it is clear that we must select a very small sample from a large data base, and that selection is obviously going to buttress our theoretical stance. This selection issue is nontrivial. The problem is how to avoid misrepresenting the data, however unintentionally.

I refer to the problem of data selection as the Bartlett Effect. In his classic work, *Remembering,* Bartlett (1932) undertook the study of systematic changes in the memory trace. One method he used, the method of serial reproduction, examined subjects' retention of folk stories selected because of their unusual content, the most famous story being *The War of the Ghosts.* One of Bartlett's predictions was that changes in the recall of this story over multiple retellings would initially distort and then convert unfamiliar content into familiar material. And indeed this is exactly what he found. However, Bartlett selected the recall protocols to illustrate his point. Bartlett viewed the protocols through the eyes of his theory. Subsequent studies that scored all recalls somewhat more objectively, although replicating the main Bartlett effect, found a great deal more variety in the recalls than Bartlett would have liked. It is particularly ironic that Bartlett was trying to prove exactly this point that we see the world through our existing knowledge and beliefs.

This problem of the theorist selecting those segments that prove his or her point is endemic in research that depends on transcripts or protocols culled from a large data base. One potential solution is illustrated in Schoenfeld's (1992) article, where he argues that the data base and scoring criteria should be made available to the field. Although we applaud this position, it is hardly practical, even in Schoenfeld's case, for it took his team a considerable period of time to analyze the 7 hours of one subject's learning sessions. One can scarcely expect the field to devote the necessary time. For archival reasons, though, Schoenfeld is right and we continue to keep all notes, audio and videotapes, and transcripts on file so that our selection bias can be checked by a later indefatigable sleuth, even though we believe that this is unlikely to occur. The influence that the introduction of archival video recordings will have on the field has hardly begun to be felt.

In order to make my task somewhat manageable, I now have observers, ethnographers, teachers, and even children indicating in their field notes when an interesting interaction occurs, so that we can transcribe "just those events of interest." This is, of course, selection before selection. But it is a nontrivial task to capture the rich social and intellectual life of a classroom with a level of analysis that would permit one to look at real conceptual change taking place over time.

LESSONS LEARNED FROM THE HISTORY OF EDUCATIONAL DESIGN EXPERIMENTS

Setting aside purely methodological issues, there is another important set of concerns that face the designer of educational interventions. There is a long history of attempts to reconstruct learning and work environments. Contemporary attempts cannot be viewed in an historical vacuum. In this section I look at lessons to be learned from some of the most famous design experiments. I refer to these as: (a) The Hawthorne Effect (Roethlisberger & Dickson, 1939)—the nature of control in multifaceted interventions, (b) The Dewey Effect—the romanticization of the process of discovery, and (c) The Reality Principle—the problem of the shelf life of successful interventions.

The Hawthorne Effect

The Hawthorne Effect, as it appears in standard texts, refers to the fact that any intervention tends to have positive effects merely because of the attention of the experimental team to the subjects' welfare. The infamous Hawthorne effect has been dogging my trail for a long time. Reciprocal teaching's success has been called "*only* a Hawthorne effect." The success of my interactive classrooms is "*merely* an example of the Hawthorne effect." Everywhere I go I can predict that someone will tell me that my results are *just* a Hawthorne effect. So just what is the Hawthorne effect? Received wisdom tells us that in an experiment conducted at the Hawthorne plant of Western Electric in the 1920s, psychologists examined the working conditions of plant workers doing repetitive tasks.

The major finding quoted in secondary sources is that irrespective of what one does to improve or degrade conditions, productivity goes up. The usual example given in secondary sources is variation in light. If light conditions improved, so did productivity; however, when light conditions were degraded, productivity also improved. The standard interpretation of these findings is that the mere presence of a research team will lead to enhanced performance because of the motivational effect of the attention received by the "subjects."

I have never taken the Hawthorne criticism of my work seriously because of the very specific nature of the improvements obtained. If I were creating a true Hawthorne effect, I would not be able to predict which performance would improve. But in fact we see a close coupling of the cognitive activities practiced and the type of improvements shown. In reciprocal teaching, only certain forms of comprehension activities increase; in the students' writing, only certain forms of structures evolve, and those are clearly predictable from the kinds of cognitive practice that the students engage in. For example, the development of hierarchical structures in the students' writing exactly paralleled the development of competence in using the file structure on the Macintosh interface. In order to find their notes on, for instance, the crested rat, students had to know that this animal is an example of animal defense mechanisms under which topic they needed to enter the file on mimicry, know that it is necessary to refer to the file on visual mimicry, and only then would they find the animal in question, one that visually mimics a skunk to defend itself. Forced to organize information into files within files within files, the children regularly traced a route through a hierarchical organization, an organization structure that was appropriated into their writing (Brown et al., in press). The very specificity of cognitive improvement (i.e., certain practiced activities result in predictable improvements in some areas and not in others) sets findings like ours apart from the Hawthorne effect. A classic example of the extreme specificity of practice and improvement is Scribner and Cole's (1981) work on the cognitive consequences of various forms of literacy among the Vai, but there are many examples of this specificity and predictability in the patterns of improvement in the literature, and when they occur they rule out simplistic recourse to a Hawthorne effect.

To satisfy my own curiosity, however, I decided to revisit the original Hawthorne work (see Roethlisberger & Dickson, 1939, for a complete overview of the multiple-year project that began in 1927). The result was illuminating. First, the famous light experiment was only one (and one of the worst documented) of approximately 35 experiments involving not only repetitive workers (relay assemblers, mica splitters) but also foremen and other decision makers. And the results were not quite as simple as one might think from reading only secondary sources.

The studies can be divided into two main types. The first concerned the physical conditions of work (light, air, hours, rest pauses, etc.) and their physical and psychological effects on productivity (fatigue, monotony). The second set concentrated on the psychological conditions of work, such as perceived status, and the effects of such activities as asking workers to serve as consultants or collaborators. To summarize briefly, the main findings were:

1. All manipulations did not result in improvement.
2. When improvements occurred, they did so under three general condi-

tions: (a) Workers *perceived* there to be improvements in the conditions being manipulated, whether or not this was so; (b) workers *perceived* the changes to be in their interest; and (c) workers *perceived* that they were in control of their own conditions of work, that is, that they were truly consultants or coinvestigators in the research endeavor. For example, the provision of rest pauses was beneficial only if the workers thought that management had their interest at heart. If the impression was that rest pauses were provided only as a Machiavellian method of increasing productivity, improvements were less apparent. The best results occurred when workers felt that they were consulted about the effects of rest pauses on both worker satisfaction and productivity.

But note that this illusion of control, or real control, is one of the things I want to happen in my classrooms. I want students to act as consultants, to be coinvestigators of their own learning (Scardamalia & Bereiter, 1983), to take charge of their own learning environment to the extent possible (Brown, 1985). On the ground that the best defense is a good offense, I argue that, redefined in this way, the Hawthorne effect is exactly what I am aiming for in my classrooms. The Hawthorne experiments have a great deal to say about worker and student responsibility and satisfaction.

The Hawthorne effect also has a great deal to say about learning in complex social environments. Consider an interesting point: Why was the intervention perceived to be a failure; why does it serve as a cautionary tale in secondary sources? Actually, only the experimenters thought it was a failure. Management was delighted with the improved esprit de corps, productivity, reduced drop-out rate, and so forth. The experimenters' feeling of failure is illustrated nicely in the following quotation (Roethlisberger & Dickson, 1939):

> The difficulty, however, went much deeper than the personal feelings of failure of the investigators. They were entertaining two incompatible points of view. On the one hand, they were trying to maintain a controlled experiment in which they could test for the effects of *single variables* while holding all other factors constant. On the other hand, they were trying to create a human situation which remained unaffected by their own activities. It became evident that in human situations not only was it practically impossible to keep all other factors constant, but trying to do so in itself introduced the biggest change of all; in other words, the investigators had not been studying an ordinary shop situation but a socially contrived situation of their own making.
>
> With this realization, the inquiry changed its character. No longer were the investigators interested in testing for the effects of single variables. In the place of a controlled experiment, they substituted the notion of a social situation which needed to be described and understood as a system of interdependent elements. (p. 185)

This reflects the major problem of trying to conduct design experiments consisting of many interwoven aspects. Components are rarely isolatable, the whole really is more than the sum of its parts. The learning effects are not even simple interactions, but highly interdependent outcomes of a complex social and cognitive intervention. And this presents a methodological headache for traditional psychology, allergic as it is to multiply confounded experiments.

To illustrate this tension between normal laboratory practice and studying learning in situ, take the example of group composition in laboratory versus classroom studies. Normal operating procedure is to select subjects either randomly or according to preexisting criteria and, if the latter, to match them closely with those who will serve in the control group. The experimenter uses standardized procedures and has control over time on task and other pertinent variables. If variables are aggregated, component analysis studies can be conducted to disaggregate the effects as part of a programmatic research effort. This was the case in the original reciprocal teaching work (Palincsar & Brown, 1984). Students were selected for the reading intervention because they met stringent criteria, their decoding rate was at grade level but they were 2 or more years behind in comprehension scores. Considerable screening was needed to generate these students. Control group students also met the criteria. The experimenter was either one of the authors or an experienced teacher carefully selected and trained; care was also taken to ensure that everyone followed the procedure correctly—time on task was carefully controlled across experimental and control groups. When reciprocal teaching was successful, the relative contributions of component strategies, procedures, materials, etc. were examined in a series of follow-up studies that teased the components apart (Brown & Palincsar, 1987, 1989).

In contrast, subsequent reciprocal teaching studies have taken place in naturally occurring groups situated in classroom settings. The students varied considerably in their ability, teachers were no longer experienced volunteers, and they had less control of time; procedural reliability became a constant problem. Control groups were difficult to engineer because of resource limitations, and even moral problems. It is difficult to expend limited resources on control settings if they involve the testing of whole classrooms, and there is the moral problem of excluding children who need the intervention. These "naturalistic" studies are inherently multiply confounded, and it would take enormous resources to unconfound them, even if this were hypothetically possible.

In the light of this problem, consider the design experiment that my research team is currently trying to engineer in the classroom. This includes effecting basic change in the role of students and teachers, modifying assessment, introducing a novel curriculum, establishing a technologically rich environment, setting up cooperative learning situations, establishing a classroom ethos where individual responsibility and group collaboration are the norm, and so forth. In short, we intervene in all aspects of the environment. Our interventions are deliberately designed to be multiply confounded. Although I was taught to avoid such messy things like the plague, I do not see an alternative. Of course one always tries to "control for" the obvious (as illustrated in Figure 4). One can effect the classroom changes without computer support, one can introduce the computer environment without attending to the curriculum, and so forth. Given the systemic nature of the environment, however, simple controls can never be entirely satisfactory; but they can provide insights into the operation of some of the major variables. Nonetheless, my major defense is that a "Hawthorne effect" is what I want: improved cognitive productivity under the control of the learners, eventually with minimal expense, and with a theoretical rationale for why things work.

The Dewey Effect

Having attempted to defend against the accusation of merely creating a Hawthorne effect, let us turn to the next most common criticism, that the work is merely a recapitulation of Dewey. After all, we have projects, discovery learning, child-centered activity, readiness to learn—what's new? Although it is undoubtedly true that the research shares philosophical commonalities with Dewey's approach, as it also does with Binet's mental orthopedics (Binet, 1909; Brown, 1985), Bruner's seminal writings about education (Bruner, 1963, 1969), and the philosophy behind the curriculum reforms of the 1960s (Schwab, 1963), there are also some subtle and not so subtle differences. I illustrate this point by considering three interrelated tenets of Dewey's (1900, 1929) faith: readiness to learn, discovery learning, and the curriculum and society.

Readiness to learn. One of the central themes of Dewey's (1929) philosophy is the notion of readiness to learn, readiness that is defined both in terms of the child's cognitive level and his or her current place in society. It has become a truism that educational practice should take into consideration the child's existing level of competence, knowledge, and interest. But Dewey seemed to imply that all education should be situated in the child's realm. Although considering the child's level is now conventional wisdom, when it is honored it is often used as justification for not teaching something, just as a superficial knowledge of Piagetian stage theory is often taken as license not to teach because a certain child has not yet reached Stage 2, is not yet ready to learn process X, Y, or Z.

Although Dewey was undoubtedly correct when he argued for diagnosis of a child's capacities, interests, and habits as a launching point for instruction, Bruner (1969) also made a telling point when he counterargued that "a point of departure is not an itinerary" (p. 116).

> It is just as mistaken to sacrifice the adult to the child as to sacrifice the child to the adult. It is sentimentalism to assume that the teaching of life can be fitted always to the child's interests just as it is empty formalism to force the child to parrot the formulas of adult society. Interests can be created and stimulated. In this sphere it is not far from the truth to say that supply creates demand, that the provocation of what is available creates response. One seeks to equip the child with deeper, more gripping, and subtler ways of knowing the world and himself. (Bruner, 1969, pp. 117–118)

My own work on readiness has been influenced by the Vygotskian notion of a Zone of Proximal Development (Brown & French, 1979; Brown & Reeve, 1987; Campione, 1989; Campione & Brown, 1990), the distance between current levels of comprehension and levels that can be achieved in collaboration with other people or powerful artifacts. Vygotsky introduced this developmental theory partially in response to his responsibilities as director of Moscow's Institute of Defectology (Vygotsky, 1978). Contradicting the prevailing notion (still alive today) that retarded children should be taught through concrete, simplified means because that is what they are "ready for," he argued that one should go beyond the current level of competence, stretch the limits, and take children to the upper boundaries of their potentiality. Vygotsky championed an introduction to the abstract, complex academic ideas that do not necessarily make contact with lived experience. Far from readiness as a notion to block off certain avenues of learning, Vygotsky argued for measures of readiness that would indicate upper levels of competence through which a child could navigate with help. Furthermore, these upper boundaries are not seen as immutable, rather they are constantly changing as the child becomes independently competent at preceding levels. This optimistic concept of ever expanding capabilities lies behind both our teaching and assessment methodologies (Brown, Campione, Webber, & McGilly, 1992; Campione, 1989). Too often Dewey's position has been interpreted in the negative, blocking sense rather than the optimistic, stretching sense.

Discovery learning. A second popular tenet of Dewey's (1929) pedagogical creed is the concept of discovery learning. The children learn best when discovering for themselves the verities of life. This notion has been fully incorporated into constructivist theories of learning, and one can surely have little critical to say about it. Discovery learning, when successful, has much to recommend it. The motivational benefits of generating knowledge cannot be overestimated, and the sense of ownership that this creates is a powerful reward. Successful discovery learning is clearly a desired feature of our communities of learners, where students are encouraged to discover, own, and share expertise.

Discovery learning is often contrasted with didactic instruction, and given that choice, I vote for discovery. There is considerable evidence that didactic teaching leads to passive learning, but by the same token, unguided discovery can be dangerous too. Children "discovering" in our biology classrooms are quite adept at inventing scientific misconceptions. For example, they readily become Lamarckians, believing that acquired characteristics of individuals are passed on and that all things exist for a purpose (teleological reasoning). They overdetermine cause, thus blinding themselves to essential notions of randomness, spontaneity, and chance.

Although it is commonplace for teachers to be called upon to foster "discovery," the role of the teacher in discovery learning classrooms is still largely uncharted. Invoking comfortable metaphors such as the teacher as coach does not tell us how and when the teacher should intervene. We know that challenging students' assumptions, providing them with counter-examples to their own rules, and so forth are good instructional ploys. But how intrusive should teachers be, when should they guide, when should they teach? When should they leave well enough alone? In short, how can teachers foster discovery and at the same time furnish guidance?

We believe that the middle ground of *guided discovery* is the most appropriate for our classrooms, but this role is difficult to maintain. Consider the position of teachers that know something that the students do not. They are in the position of making a judgment call about whether to intervene or not: How long should they let the students flounder? They must decide whether the problem surrounds an important principle that needs work or involves only a trivial error that they can let pass for now. Consider the case of teachers who do not know the answer, or who may share the students' puzzlement or misconception. In this case they are first required to recognize this fact (which might not be easy) and, after admitting puzzlement or confusion, find ways to remedy it, for example by seeking help. This is not an easy role for many teachers; it demands competence and confidence. The provision to our classrooms of an electronic mail system that links the teachers and students to a community of scholars (biologists, etc.) helps teachers handle the lack of knowledge problem.

Guided learning is easier to talk about than do. It takes clinical judgment to know when to intervene. Successful teachers must engage continually in on-line diagnosis of student understanding. They must be sensitive to overlapping current zones of proximal development, where students are ripe for new learning. Guided discovery places a great deal of responsibility in the hands of teachers, who must model, foster, and guide the "discovery" process into forms of disciplined inquiry that would not be reached without expert guidance.

The curriculum and society. Finally, Dewey (1929) stressed the need to situate curriculum activity in the lives of children. Curricula should reflect the child's lived experience and provide continuity with the family and community life. The idea was that if children were involved in projects where they ran farms, built treehouses, and so forth, they would learn not only mathematics, for example, but they would understand better the purpose of mathematics. Project type interventions were popular for a while until the inevitable conservative backlash (Cremin, 1961) and we are seeing a great deal of renewed interest today. Indeed, Dewey himself denounced a great deal of project-type work purportedly conducted in the spirit of his Chicago school innovation (Dewey, 1931; Kliebard, 1987).

The spirit of Dewey's position was intended to counteract the isolation of much of school learning from the familiar habits of childhood on the one hand, and adult occupations on the other (Cole & Bruner, 1971); however, there is another side to the story. Schools came into existence to expose children to knowledge outside the realm of lived experience, an exposure rarely possible in preliterate society. As Bruner (1969) argued, to understand something is to relinquish prior ways of thinking, and he is skeptical about the practice of situating learning exclusively in the child's lived experience and social activities. For example, he argued that:

> The significance of the concept of commutativity in mathematics does not derive from the social insight that two houses with fourteen people in each is not the same as fourteen houses with two people in each. Rather it inheres in the power of the idea to create a way of thinking about number that is lithe and beautiful and immensely generative. . . . Without the idea of commutativity, algebra would be impossible. If set theory . . . had to be justified in terms of its relation to immediate experience and social life, it would not be worth teaching. Yet set theory lays a foundation for the understanding of order and number that could never be achieved with the social arithmetic of interest rates and bales of hay at so much per bale. (Bruner, 1969, p. 121)

"Lithe and beautiful and immensely generative" ideas are assuredly what one is searching for as a basis for a curriculum. And it is unreasonable to expect children to reinvent pivotal ideas for themselves. Providing expert guidance in the form of teachers, or books and other artifacts, is one of the prime responsibilities of schooling. Immensely generative ideas may be few, and the idea behind education is to point children in the right direction so that they may discover and rediscover these ideas continuously. This notion of a spiraling curriculum (Bruner, 1969) is one that we try to embody in our classrooms. Central themes are introduced by the teacher early and revisited often. In the environmental science classroom, these themes might include notions of interdependence and balance, competition and cooperation, and adaptation that are central to an understanding of ecosystems.

Coupled with these recurrent themes are habits of mind by which children are encouraged to extrapolate, refine, and use these underlying themes so that they can discover new commonalties for themselves. As Bruner argued (1969), education:

> should be an invitation to generalize, to extrapolate, to make a tentative intuitive leap, even to build a tentative theory. The leap from mere learning to using what one has learned in thinking is an essential step in the use of the mind. Indeed, plausible guessing, the use of the heuristic hunch, the best employment of necessarily insufficient evidence—these are the activities in which the child needs practice and guidance. They are among the great antidotes to passivity. (p. 124)

Note that this is not untrammeled discovery learning, but learning clearly under the expert guidance of a gifted teacher.

Although recognizing an intellectual debt to Dewey, it is also true that a great deal of theoretical and methodological work is still needed if we are to render these ideal educational types into anything close to reality. Collins (in press) argued that educational theorists (Plato, Rousseau, Dewey, Bruner, and Illich) have addressed themselves to the process of education for many years. The contemporary agenda is for experimentalists to fulfill this legacy with concrete design experiments. Such experiments must provide a level of description that would afford the opportunity to uncover mechanisms of learning that are not captured in preexisting theoretical descriptions. So in response to the criticism, "It's all Dewey, what's new?" the answer is, nothing and everything.

The Reality Principle

The final critical issue concerning educational reform I address is whether or not it has any lasting effect. In the light of the poor long-term track record of the laboratory schools of Binet, Dewey, and others, one must consider the shelf life of educational interventions warily. Received wisdom tells us that such innovations as the Dewey school and the science curriculum reforms of the 1960s were experiments that were tried and failed. But it is not clear what was meant by tried or failed, nor is it clear what the criteria were against which success or failure is measured (Dow, 1991; Shymansky, 1989). Historians tend to be pessimistic about educational reform (Cuban, 1984, 1986, 1990), an attitude of "la plus ça change" (Cohen, 1989) prevails. The argument is that successful interventions are a chimera or at least are extremely fleeting and fragile, not readily transportable to settings outside the innovator's control. Because of this skepticism, it is extremely important for the design experimenter to consider dissemination issues. It is not sufficient to argue that a reasonable end point is an existence proof, although this is indeed an important first step.

In this light, the long-term track record of reciprocal teaching is encouraging; the method enjoys widespread dissemination compared with other innovative instructional designs with which I have been involved. I believe that the reasons for this are (a) it looks deceptively easy to implement, (b) the term *reciprocal teaching* has been picked up by researchers, teachers, and textbook publishers and has become part of the discourse of the reading community, and (c) most important, the procedure slots neatly into a hallowed classroom niche (Cuban, 1984). Reading group has a long-standing place in the school day—teachers are used to arranging their classes to accommodate reading group; all (all!) that is needed is to redefine the activities that take place in a socially sanctioned niche.

The prognosis for the widespread dissemination of communities of learning is pessimistic. The desired participant structures of this program (Brown et al., in press) would require fundamental changes in the roles of both students and teachers, disrupting "practice as usual," and we know that historically teachers have been resistant to such disruptions.

To say the least, it is a cautionary note for contemporary designers that Dewey (1901) a century ago warned that educational reform would not be easy to engineer. Dewey's description of cycles of innovation and resistance sounds uncannily like Cuban's (1984, 1990) contemporary Cassandra bulletins. First comes unrest concerning the schools and how they operate, followed by fervent claims and promises from reformers. Intensive research by the converted is then carried out in a small set of classrooms rich with human and, today, technological resources. "The victory is won and everybody—unless it is some already overburdened and distracted teacher—congratulates everybody else that such advanced steps can be taken" (Dewey, 1901, p. 334). But then come the frustrated attempts by ordinary teachers to adopt the new methods in the absence of support, followed by the inevitable decline in use, and the eventual abandonment of the program. As Dewey argued, "within a short time, complaints are heard that children do not read as well," "or a public outcry calls for the reforms to be rescinded in favor of the status quo." One major question facing contemporary designers is how to avoid repeating the Cuban–Dewey cycle: exhilaration, followed by scientific credibility, followed by disappointment and blame.

I see the problem as analogous to the alpha, beta, and gamma phases of software development. The alpha, or developmental, phase is under the control of the advocate, and by definition it must work for there to be any later phases. It works, though, under ideal supportive conditions. Next comes the beta phase, tryouts at carefully chosen sites with less, but still considerable, support. Critical is the gamma stage, widespread adoption with minimal support. If this stage is not attempted, the shelf life of any intervention must be called into question.

My classroom research agenda now is far less that of the frustrated

Hawthorne experimenter trying to unconfound variables for explanatory power, although I still care about control issues. My agenda is more like that of a designer or engineer. I need to unconfound variables, not only for theoretical clarity, but also so that necessary and sufficient aspects of the intervention can be disseminated. The question becomes, what are the absolutely essential features that must be in place to cause change under conditions that one can reasonably hope to exist in normal school settings?

In order to effect this last stage, I again need additional expertise, more methods if you like, that were not part of my training. I need to know a great deal more about school restructuring, teacher training and support, and teachers as researchers. I need to use ethnographic methodologies to study which aspects of the program are readily adopted and which are resisted. I need to know the conditions favorable for adoption. I need to study the sociology of dissemination. I need to know about public policy issues, and so forth. Again changes in theory and practice demand concomitant advances in methodology for the conduct, documentation, and dissemination of research.

SEX, LIES, AND VIDEOTAPES

I couch my final comments on methodology in terms of this facetious but not entirely inappropriate subtitle. Designers of successful classroom interventions must make sure that they are engaging enough to seduce children into the world of learning, hence "sex" in the title. Once ensnared, it may be possible to guide students toward the intrinsic rewards that follow from self-initiated disciplined inquiry. As a design scientist, it is necessary to tease apart the major features of enticing learning environments: the role of teachers, students, and researchers; the actual contribution of curricula and computer support; methods by which distributed expertise and shared meaning are engineered, and so forth. There is a constant tension between designing an exciting classroom for happy campers and maintaining research standards of control and prediction.

The excitement of teachers and students is palpable when it all works, when they experience what one of my teachers calls "golden moments." However, in reporting these moments, it is important to realize that they are selected from a stream of events that are not so encouraging. One must separate the gold from the dross. But how does one represent the gold–dross ratio? Hence "lies" in the title; there is a tendency to romanticize research of this nature and rest claims of success on a few engaging anecdotes or particularly exciting transcripts. One of the major methodological problems is to establish means of conveying not only the selective and not necessarily representative, but also the more important general, reliable, and repeatable.

Finally, why "videotapes." Several years ago I was writing a report for a government agency about major advances in developmental psychology in the preceding decade, and I wrote to colleagues asking for their candidates. One colleague replied with a single word: videotapes. My first reaction was that he was being typically facetious, but on further thought, I realized that this response was profound. The advantage that modern designers of classroom innovations have is that they make videotape records for archival purposes. Dewey, Binet, and other earlier designers had to depend on the selective reporting of the innovator and his disciples. But today videotape records serve many functions: They can be reviewed in the light of changing perspectives. They can be used to look back on the performance of children or groups of children who are interesting examples of a certain learning phenomenon and so forth. Tapes are invaluable for documenting conceptual change, in teachers as well as students, and they provide a common data base for discussion and reflective action on the part of teachers and researchers. Today videotape records of successful teacher interventions, peer tutoring, discussions, plans, insights, and generative ideas can be preserved so that they can be analyzed subsequently by those with radically different theoretical spectacles. As Schoenfeld (1992) argued in this issue, these records can be made available to the field, available for continual reanalysis as new and more powerful theories of learning and concomitant methodologies emerge.

CONCLUSION

In this article, I have described my current attempts to engage in design experiments intended to transform classrooms from academic work factories to learning environments that encourage reflective practice among students, teachers, and researchers. I traced briefly the history of my own progression from the study of laboratory learning to classroom observations and experimentation. The need for new and complex methodologies to capture the systemic nature of learning, teaching, and assessment was described, as was the need to consider the history of prior attempts to reorganize school and work environments.

ACKNOWLEDGMENTS

An earlier version of this paper was presented at the meetings of the American Educational Research Association, Chicago, 1991 in a symposium entitled "On paradigms and methods: What do you do when the ones you know don't do what you want them to?"

The preparation of this paper was supported by grants from the James S. McDonnell and Andrew W. Mellon Foundations, and from the Evelyn Lois Corey Research Fund.

I express my appreciation to my colleagues, notably Joseph Campione and Annemarie Palincsar, who are jointly responsible for a great deal of the work described in this article, and to Alan Schoenfeld who provided important feedback on an early draft of this article.

REFERENCES

Aronson, E. (1978). *The jigsaw classroom.* Beverly Hills, CA: Sage.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In R. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89-195). New York: Academic.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge: Cambridge University Press.

Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 361-392). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Binet, A. (1903). *L' Ètude experimentalè de l'intelligence* [The experimental study of intelligence]. Paris: Schleicher Frères.

Binet, A. (1909). *Les idees modernes sur les infants* [Modern ideas about children]. Paris: Ernest Flammarion.

Brown, A. L. (1974). The role of strategic behavior in retardate memory. In N. R. Ellis (Ed.), *International review of research in mental retardation* (Vol. 7, pp. 55-111). New York: Academic.

Brown, A. L. (1975). The development of memory: Knowing, knowing about knowing, and knowing how to know. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 10, pp. 103-152). New York: Academic.

Brown, A. L. (1977). Development, schooling and the acquisition of knowing about knowledge. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 241-253). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 1, pp. 77-165). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Brown, A. L. (1985). Mental orthopedics: A conversation with Alfred Binet. In S. Chipman, J. Segal, & R. Glaser (Eds.), *Thinking and learning skills: Current research and open questions* (Vol. 2, pp. 319-337). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Brown, A. L., Ash, D., Rutherford, M., Nakagawa, K., Gordon, A., & Campione, J. C. (in press). Distributed expertise in the classroom. In G. Salomon (Ed.), *Distributed cognitions.* New York: Cambridge University Press.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology (4th ed.): Cognitive development* (Vol. 3, pp. 77-166). New York: Wiley.

Brown, A. L., & Campione, J. C. (1978). Permissible inferences from cognitive training studies in developmental research. *Quarterly Newsletter of the Institute for Comparative Human Behavior, 2*(3), 46-53.

Brown, A. L., & Campione, J. C. (1979). The effects of knowledge and experience on the formation of retrieval plans for studying from texts. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical Aspects of Memory* (pp. 378-384). London: Academic.

Brown, A. L., & Campione, J. C. (1986). Psychological theory and the study of learning disabilities. *American Psychologist, 41*(10), 1059-1068.

Brown, A. L., & Campione, J. C. (1990). Communities of learning and thinking, or a context by any other name. *Human Development, 21,* 108-125.

Brown, A. L., & Campione, J. C. (in press). Restructuring grade school learning environments to promote scientific literacy. In *Restructuring learning: Analysis and recommendations of the Council of Chief State School Officers.* Washington, DC: Council of Chief State School Officers.

Brown, A. L., Campione, J. C., Reeve, R. A., Ferrara, R. A., & Palincsar, A. S. (1991). Interactive learning and individual understanding: The case of reading and mathematics. In L. T. Landsmann (Ed.), *Culture, schooling and psychological development* (pp. 136-170). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Brown, A. L., Campione, J. C., Webber, L. S., & McGilly, K. (1992). Interactive learning environments—a new look at assessment and instruction. In B. R. Gifford & M. C. O'Connor (Eds.), *Future assessment: Changing views of aptitude, achievement and instruction* (pp. 121-213). Boston: Academic.

Brown, A. L., & French, L. A. (1979). The zone of potential development: Implications for intelligence testing in the year 2000. *Intelligence, 3,* 253-271.

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology, 20,* 493-523.

Brown, A. L., & Palincsar, A. S. (1982). Inducing strategic learning from texts by means of informed, self-control training. *Topics in Learning and Learning Disabilities, 2*(1), 1-17.

Brown, A. L., & Palincsar, A. S. (1987). Reciprocal teaching of comprehension strategies: A natural history of one program for enhancing learning. In J. D. Day & J. G. Borkowski (Eds.), *Intelligence and exceptionality: New directions for theory, assessment, and instructional practice* (pp. 81-132). Norwood, NJ: Ablex.

Brown, A. L., & Palincsar, A. S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 393-451). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Brown, A. L., & Reeve, R. A. (1987). Bandwidths of competence: The role of supportive contexts in learning and development. In L. S. Liben (Ed.), *Development and learning: Conflict or congruence?* (pp. 173-223). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Bruner, J. S. (1963). *The process of education.* Cambridge, MA: Harvard University Press.

Bruner, J. S. (1969). *On knowing: Essays for the left hand.* Cambridge, MA: Harvard University Press.

Campione, J. C. (1989). Assisted assessment: A taxonomy of approaches and an outline of strengths and weaknesses. *Journal of Learning Disabilities, 22,* 151-165.

Campione, J. C., & Brown, A. L. (1987). Linking dynamic assessment with school achievement. In C. S. Lidz (Ed.), *Dynamic assessment* (pp. 82-115). New York: Guilford.

Campione, J. C., & Brown, A. L. (1990). Guided learning and transfer: Implications for approaches to assessment. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 141-172). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Campione, J. C., Brown, A. L., & Connell, M. L. (1988). Metacognition: On the importance of understanding what you are doing. In R. I. Charles & E. A. Silver (Eds.), *Research agenda for mathematics education: The teaching and assessing of mathematical problem solving* (pp. 93-114). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Campione, J. C., Brown, A. L., & Jay, M. (in press). Computers in a community of learners.

In E. De Corte, M. Linn, H. Mandl, & L. Verschaffel (Eds.), *Computer-based learning environments and problem solving.* (NATO ASI Series. F: Computer and Systems Series). Berlin: Springer.

Campione, J. C., Rutherford, M., Gordon, A., & Brown, A. L. (in press). Communities of learners: Zones of development for those at risk. In N. C. Jordan & J. Goldsmith-Phillips (Eds.), *Learning disabilities: New directions for assessment and intervention.* Needham Heights, MA: Allyn and Bacon.

Chi, M. T. H. (1976). Short-term memory limitations in children: Capacity or processing deficits? *Memory and Cognition, 4,* 559-572.

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73-96). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cohen, D. (1989). Teaching practice: Plus ça change . . . In P. Jackson (Ed.), *Contributing to educational change: Perspectives on research and practice* (pp. 27–84). Berkeley, CA: McCutchan. (Also published by the National Center for Research on Teacher Education. Michigan: Michigan State University, 88–3, September 1988.)

Cole, M., & Bruner, J. S. (1971). Cultural differences and inferences about psychological processes. *American Psychologist, 26,* 867–876.

Collins, A. (in press). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology.* New York: Springer-Verlag.

Collins, A., & Stevens, A. (1982). Goals and strategies of inquiry teachers. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 2, pp. 65–119). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cremin, L. A. (1961). *The transformation of the school: Progressivism in American education 1876–1957.* New York: Knopf.

Cuban, L. (1984). *How teachers taught: Constancy and change in American classrooms, 1890–1980.* New York: Longman.

Cuban, L. (1986). *Teachers and machines: The classroom of technology since 1920.* New York: Teachers College Press.

Cuban, L. (1990). Reforming again, again and again. *Educational Researcher, 13,* 3-13.

Darwin, C. (1877). A biographical sketch of an infant. *Mind, 2,* 286–294.

DeLoache, J. S., Brown, A. L., & Kane, M. J. (1985, April). *Young children's error correction procedures in manipulative play.* Paper presented at the meeting of the Society for Research in Child Development, Toronto.

Dewey, J. (1900). *The school and society.* Chicago: University of Chicago Press.

Dewey, J. (1901). The situation as regards the course of study. *Journal of Proceedings and Addresses of the Fortieth Annual Meeting of the National Education Association,* 332–348.

Dewey, J. (1929). *My pedagogical creed.* Washington, DC: Progressive Education Association.

Dewey, J. (1931). *The way out of educational confusion.* Cambridge, MA: Harvard University Press.

Dow, P. (1991). *Schoolhouse politics.* Cambridge, MA: Harvard University Press.

Fish, S. (1980). *Is there a text in this class? The authority of interpretive communities.* Cambridge, MA: Harvard University Press.

Flavell, J. H. (1971). First discussant's comments: What is memory development the development of? *Human Development, 14,* 272–278.

Flavell, J. H., & Wellman, H. M. (1977). Metamemory. In R. V. Kail, Jr. & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 3-33). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155–170.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony

(Eds.), *Similarity and analogical reasoning.* Cambridge: Cambridge University Press.
Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12,* 306-355.
Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology, 40,* 631-666.
Hatano, G., & Inagaki, K. (1987). Everyday biology and school biology: How do they interact? *The Newsletter of the Laboratory of Comparative Human Cognition, 9,* 120-128.
Karmiloff-Smith, A., & Inhelder, B. (1974). If you want to get ahead, get a theory. *Cognition, 3,* 195-212.
Kessen, W. (1965). *The child.* New York: Wiley.
Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363-394.
Kliebard, H. M. (1987). *The struggle for the American curriculum 1893-1958.* New York: Routledge.
Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills.* New York: Academic.
Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.
Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition and Instruction, 1*(2), 117-175.
Palincsar, A. S., Brown, A. L., & Campione, J. C. (1990, April). *Discourse as a mechanism for acquiring process and knowledge.* Paper presented at the American Educational Research Association Conference, Boston.
Piaget, J. (1952). *The origins of intelligence in children.* New York: International Universities Press.
Reeve, R. A., Gordon, A., Campione, J. C., & Brown, A. L. (1990, April). *Enhancing and predicting the math performance of elementary school children using reciprocal teaching and dynamic assessment.* Paper presented at the American Educational Research Association Conference, Boston.
Roethlisberger, F. S., & Dickson, W. J. (1939). *Management and the worker.* Cambridge, MA: Harvard University Press.
Scardamalia, M., & Bereiter, C. (1983). Child as coinvestigator: Helping children gain insight into their own mental processes. In S. Paris, G. Olson, & H. Stevenson (Eds.), *Learning and motivation in the classroom* (pp. 61-82). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Scardamalia, M., & Bereiter, C. (1991). Higher levels of agency for children in knowledge building: A challenge for the design of new knowledge media. *The Journal of Learning Sciences, 1,* 37-68.
Schoenfeld, A. H. (1985). *Mathematical problem solving.* New York: Academic.
Schwab, J. J. (1963). *Biology teacher's handbook.* New York: Wiley.
Scribner, S., & Cole, M. (1981). *The psychology of literacy.* Cambridge, MA: Harvard University Press.
Shymansky, J. A. (1989). About ESS, SCIS, and SPAP. *Science and Children, 7*(26), 33-35.
Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology, 21,* 437-484.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. and Trans.). Cambridge, MA: Harvard University Press.
Wellman, H. M. (1983). Metamemory revisited. In M. Chi (Ed.), What is memory development the development of? A look after a decade. *Human Development* (pp. 31-51). Basel, Switzerland: S. Karger.
Whitehead, A. N. (1916). *The aims of education.* Address to the British Mathematical Society, Manchester, England.