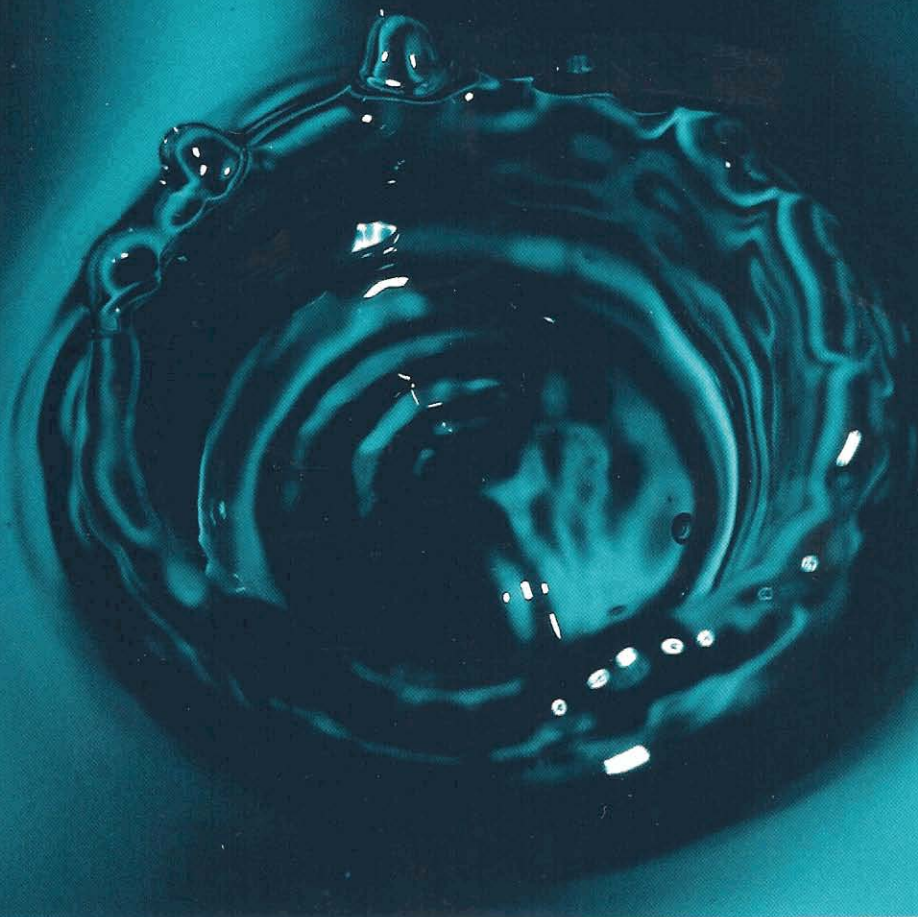


Edited by
Mark A. Conchas • Robert J. Sternberg



Translating Theory and Research Into Educational Practice

*Developments in Content Domains,
Large-Scale Reform, and Intellectual Capacity*

Learning to Do Educational Research¹

Robert C. Calfee
University of California, Riverside

Roxanne Greitz Miller
Chapman University

Kim Norman
California State University, Fullerton

Kathy Wilson and Guy Trainin
University of Nebraska, Lincoln

In this chapter, we present an account of selected research and development programs in the literacy arena conducted over the last decade under the direction of R. C. Calfee in collaboration with his coauthors. Three specific projects provide context and empirical findings for approaching the issue of translation of research into practice. Following are brief sketches of each project, laying out the framework and research strategy along with goals for practice. All three projects explore professional development strategies designed to promote fundamental change in teachers' knowledge, beliefs, and practices consistent with current sociocognitive learning theories. A concomitant concept has emerged during the work: the influence of organizational and contextual factors in generalizing the initiation and sustainability of the core concept. Following brief sketches of the three projects, we explore three themes emerging from our experiences in translating research into practice and then address the four questions posed by the editors.

¹Support provided by Spencer 199900046, IERI 9979834.

THREE PROJECTS

READ-Plus

The umbrella project for this collection, READ-Plus, is a comprehensive and developmental literacy program incorporating four curriculum elements (decoding, vocabulary, narrative, exposition) within a constructivist instructional design also comprising four elements (connect, organize, reflect, extend); the early history of the project is recounted by Calfee and Patrick (1995). The project began in 1980 in response to the request by Graystone Elementary School in San Jose, California, for a review of their literacy program and suggestions for choosing a new basal series. R. C. Calfee and a team of Stanford graduate students met with the faculty during a summer workshop to discuss the implications of social-cognitive principles for practice, focusing on teacher knowledge and beliefs. Although the workshop had been requested by the teachers, the initial response was rather negative: "Too much theory! When will you tell us what to do? Why worry about memory and attention—we need to teach reading." The workshop was indeed rough around the edges; it was the team's first serious foray into a practical setting, and the materials were more suited to a doctoral seminar than a practitioner workshop. Our model of collaboration was that the Stanford team would present theory and research, and the Graystone team would provide the "so what" based on their extensive experience. The day was saved by two people: Jean Funderburg, the principal, and Doris Dillon, an outstanding kindergarten teacher. Funderburg offered low-key but persistent encouragement of the faculty following each session, and Dillon expressed enthusiasm for a framework that made sense of what she did in her classroom. By the end of the workshop, the teachers had committed to trying out the framework during the following school year. They would continue with the same basal series but prioritize some curriculum segments (e.g., stories with a solid narrative structure) and drop others (rote learning of vocabulary) while adopting an instructional strategy quite unlike the initiate-respond-evaluate (IRE) embedded in the Teacher's Manual (Cazden, 1988). The Stanford team developed *THE BOOK for the Reading Teacher* (Calfee & Associates, 1981) as a resource guide including sample lessons and assessments but in no way a replacement for the teachers manual.

The Graystone project affected major elements of curriculum, instruction, assessment, and organizational patterns over what became a multiyear collaboration. Evidence showed substantial improvements in student achievement on state indicators as well as on classroom-based assessments, particularly in the areas of writing and motivation. The impact on teachers was, from our perspective, the more important outcome. The project even-

tually came to an end as a school-wide endeavor; for 3 years in a row, Funderburg was informed by the district prior to the beginning of the school year that she would be returned to a classroom assignment and a new principal assigned and then at the last hour asked to remain as principal. After the third such experience, she took an extended leave to enter Stanford's doctoral program. Individual teachers, despite shifts in grade and school assignments, appeared to undergo fundamental changes, some of which we describe following.

Support from the Stanford team was largely voluntary and incidental. The project had little grant support, but as the years progressed, the model, which was eventually christened READ-Plus, emerged as a platform for conducting a variety of in situ investigations of literacy acquisition with external funding. When implemented as a school-wide strategy, the READ-Plus agenda entailed substantial changes in philosophy and practice. For practical reasons, we moved programmatically to explore two extensions, "Word Work" and "Reading and Writing About Science," described following. Limiting the agenda to specific curriculum components and selected grades proved more palatable, especially for schools serving large proportions of students with special needs and overwhelmed with a plethora of disconnected projects. The focal elements in these two projects—Decoding-Spelling and Exposition—were especially challenging for teachers, so it made sense to focus our research activities on these elements. To be sure, this approach brought some difficulties. READ-Plus transformed a school's contextual environment in significant ways, easing the way for experimentation across the board. A more limited agenda can foster a "band-aid" perspective—during certain times during the day and week, teacher and class engage in "something different," but otherwise it is business as usual; some teachers participate in the "special" program with little organizational impact. The focal strategy entails benefits and costs.

WordWork

The decoding/spelling component of Read-Plus, WordWork, provides kindergarten through second-grade teachers an elaborated framework for effective instruction in phonemic awareness and phonics (Calfee, 1998). Blocks of 2 weeks with short lessons spiral through a small number of highly productive letter-sound correspondences from the Anglo-Saxon layer of English, the primary source of the spelling patterns that students encounter in the primary grades. This focus on selected correspondences contrasts with current practice in which hundreds of objectives fill the phonics curriculum with weird patterns like "weird." The WordWork rationale emphasizes a streamlined and efficient phonics curriculum affording students rapid access to a variety of texts and expression in readable writing. The curriculum is sup-

ported by CORE (connect, organize, reflect, extend), the READ-Plus instructional model (Chambliss & Calfee, 1998; Miller & Calfee, 2004) we describe following. Lesson blocks incorporate the tenets of social cognition; students work in small groups actively exploring sounds and letters.

Phonemic awareness appears early in WordWork, but in ways that differ sharply with current practice. A single lesson block focuses on articulation, the production of sounds in the oral cavity. Based on the motor theory of speech perception (Calfee & Norman, 1999), students study the features of consonant production such as the contrast between popping and hissy sounds (plosives and fricatives for linguists). Organizing the consonants into categories or "chunks" provides a technical vocabulary for students (and the teacher) to think and talk about letters, sounds, and words across a range of new and unfamiliar words.

After establishing a small repertoire of consonants, the teacher immediately leads students to the phonics stage in which they connect the consonant sounds with symbols (letters). The strategy emphasizes the functional role of vowels; students learn to "make words" using vowels as "glue letters." The word *tap* can be built when "you touch your tongue to the roof of your mouth, then glue in the 'aaaa' sound and blow through your lips." WordWork emphasizes the *metaphonic* principle; students discuss why they spell a word in a particular way. For example, a student may explain that "'tape' needs the final-e as a 'buddy,' so that 'a' can 'say its name.'" Reflective talk based on an explicit technical vocabulary undergirds conceptual understanding of English orthography. Metaphonic talk is active, explicit, and social; conducted in teacher-guided, small-group settings, supported by "hands-on" devices such as letter tiles, allowing students to learn from peers, consistent with Vygotskian principles.

Reading and Writing About Science (RWS)

The RWS project centers around the Vocabulary and Exposition elements from READ-Plus. The curriculum integrates reading and writing instruction within the content area of science to improve students' reading comprehension and expository writing skills in the upper elementary and middle school grades (Grades 4–6). A significant undercurrent is the development of academic language encompassing all forms of spoken and written language (Heath, 1983). The RWS curriculum highlights the Read-Write Cycle, a variation of the CORE framework (Calfee & Miller, in press; Chambliss & Calfee, 1998). The Cycle reflects the essential connections between reading comprehension and writing performance for the expository genre.

Three principles are foundational for the RWS project. First, integration of reading and writing in instruction is key to improving students' reading comprehension and writing skills (Nelson & Calfee, 1998; Sperling & Freed-

man, 2002; Tierney & Shanahan, 1991). Second, all students, but particularly those who struggle with writing, benefit from explicit instruction in cognitive and sociocognitive strategies in reading, writing, and problem solving (Carr & Ogle, 1987; Lapp, Flood, & Hoffman, 1996; Prain & Hand, 1996). Third, incorporating research-based techniques into classroom practice depends more on teachers than on programs (Robinson, 1998). Hence, the emphasis is on instructional strategies that are *efficient*, not requiring additional enormous amounts of time to implement in the classroom; *effective*, incorporating methods that can be applied to a broad range of texts, grade levels, and subject areas; and *adaptable*, allowing the teacher to employ the strategies across a broad range of student achievement levels and interests.

In this chapter, we lay out three themes that capture our experiences during these research activities and then respond to the four questions presented by the editors. The themes are designed to provide coherence to the responses, so the reader will encounter some intentional redundancy.

THREE THEMES

Scientists learn methods for conducting research, and then they perform investigations using these techniques. This image of the scientific enterprise is promulgated from the early grades onward, including college and in some instances, even graduate study. Our position departs from this image to suggest that learning is an essential element in the research process, especially when the aim is to connect basic investigations with practical application. The three themes address this connection: (a) research as complex learning, (b) research valid for applied outcomes, and (c) research on the application of research to practice. The themes will not proceed in a linear sequence. Rather, they attempt to capture the interactive nature of educational research. The three projects sketched previously illustrate the concepts.

Research as a Complex Learning Task

The first theme centers on the view of scientific research as a learning task: a form of complex problem solving, a human enterprise, an ongoing effort to understand our existence. Science relies on shared beliefs and values to be sure, but these are scarcely fixed for all time. Rather, scientists are distinguished by the capacity to advance knowledge through concepts, methods, and interpretive techniques that undergo continual change both incremental and fundamental (Kuhn, 1970; Phillips, 2000). Critical analysis is an important part of the scientific process; can conceptual and empirical results stand up against systematic challenges (Kincaid, 1996; Stanovich, 2003)? Generalizability is a critical consideration. Newtonian principles operate

within certain boundary conditions but are not universal. This constraint does not lessen the value of the Newtonian principles, but illustrates the importance of establishing the limits of generalizability.

In educational research, the goals are to understand and influence significant social practices that are inherently complicated, dynamic, and changeable. Social policy (and hence, politics) is part of the equation. The situation may resemble arenas such as environment and transportation but seems more intense where children are involved. A technique or program is advanced as a remedy to an educational need, with accompanying conceptual and empirical support. For policymakers and practitioners, the obvious response is "go with it!" For the researcher, the obvious question is, "When and where will it work?"

The generalizability of educational research is obviously challenged by differences among people and contexts, but time and space also matter. Valid designs must account for these variations: people and contexts—the teacher in New York City confronts a setting quite unlike her counterpart in the Four Corners region: time—student learning takes place over lengthy spans as does student development. The same can be said about teachers. Historical trends also matter; consider the educational terrains from 1950 through 2025: space—a facet found in research designs idiosyncratically if at all. Organizational layers are part of this dimension: classroom, school, district, region, state, and nation (Bryk & Raudenbush, 1992). Geography offers another lens, both physical and social: from amber waves of grain and purple mountain majesties to inner city ghettos and rural outposts. Whether a principle or program—the effectiveness of direct instruction or the benefits of whole language—data-based claims must respond to critics who ask for evidence that scientific findings apply to "us in the here and now."

The point is not that education is infinitely complex but that pinning down the critical constancies in the midst of substantial diversities should be taken seriously (Calfee & Nelson-Barber, 1991). What seems to work in one setting may disappear in another context, hence, the common complaint that research findings may support almost any side of any issue. For example, does class size matter? The answer is clearly "It depends." This answer does not waffle; the research challenge is to pin down the dependencies. For a variety of reasons, educational research has been relatively unsuccessful in dealing with generalizability; even the popular meta-analytic strategies designed to identify dependencies leave much to be desired. The situation is similar in fields such as medicine and engineering despite their more exalted position (e.g., Marshall, 2003; Sanders, 2003). The public tends to overlook or forgive "failures" such as hormone replacement therapy or the freeway failures during the Loma Prieta earthquake. When a doctor experiments to help a patient experiencing a complex syndrome or an architect wrestles with a persistent roof leak, the public does not conclude that research in medicine and engineering is worth-

less. Rather, they recognize (albeit with understandable frustration) that practical application of research is a learning experience dependent on research findings, practitioner expertise, and collaboration among practitioners and researchers.

Research That Is Up to the Demands of Practice

The second theme centers primarily around issues related to the quality of educational investigations. To meet the considerable challenges of practical applications, educational research must meet high standards of scientific inquiry. This topic has garnered considerable attention in the past several years including the superb National Review Council (NRC) review by Shavelson and Towne (2002; also Natriello, 2004; Shavelson, Phillips, Towne, & Feuer, 2003), along with legislative mandates spelling out appropriate methodologies for establishing valid knowledge (Eisenhart & Towne, 2003).

Our first point under this theme is the importance of establishing a conceptual framework as a foundation; sound design, methodology, and analysis are more likely when based on a coherent framework (Chambliss & Calfee, 2002). To be sure, education does not possess the theoretical underpinnings found in some other professional arenas. More to the point, although “conceptual framework” frequently appears in discussions of research strategy, it is difficult to find explicit mention in methods textbooks. Indeed, the usual strategy remains the null-hypothesis approach from a half century ago.

A second point about quality centers on methodological adequacy. Generally speaking, multiple methods that triangulate on the problem would appear to be preferred. Debate since the 1970s has centered around “the one right method,” often replacing reasoned argument with rhetoric (Phillips, 2000). Experienced researchers are often classified as either quantitative or qualitative, which then serves as the model for doctoral students when they approach the dissertation. Methodological pendulum swings have been substantial in recent decades. Half a century ago, the prevailing model for educational research relied on experimental designs using quantitative outcomes; the target treatment was compared with a control using pretests and posttests to assess the statistical trustworthiness of the findings. The shortcomings of this design are numerous. Sir R. A. Fisher (1935), the originator of experimental design, opens his first volume by conceptualizing a practical problem in agriculture with a four-factor design as the basis for adequate control, a technical requirement for validity that is largely absent from today’s discussions (but cf. Anderson, 2001).

The 1970s saw a major shift from the previous paradigm. Funding priorities from the National Institute of Education began to emphasize qualitative methods. Such methods had always been part of the palette along with

descriptive and correlational studies and case studies. More recently, policymakers have returned to the experimental-control "gold standard," which is promoted as the foundation for cause-effect attributions (Reyna, in press). This claim is questioned by many philosophers of science, partly because experimental-control designs seldom suffice for such attributions and because other methods are generally required to inform and support causal claims (Kincaid, 1996). As Pearson (in press) notes, randomized field experiments in medicine and elsewhere depend on substantial background research, both fundamental and applied.

A third point that has emerged from our experiences centers around generalizability methods originally developed by Cronbach, Gleser, Nanda, and Rajaratnam (1972) to extend the concept of test reliability. Cronbach (1957, 1975) also reflected on generalizability as a unifying methodology for the "two disciplines of psychology." The idea was to view investigations as "tests" subject to the same requirements of reliability and validity. In essence, generalizability places on the researcher the task of identifying the facets that influence a phenomenon, both directly and through interactions. To be sure, some investigators eschew generalizability, suggesting that their findings cannot or should not be extended to other situations. Such claims of uniqueness seem to us more appropriate to art than to science, both having value in their own right.

The point is that whether advocating quantitative or qualitative approaches, experimental or descriptive techniques, all those involved in the research enterprise—investigators, practitioners, and policymakers—can benefit by reflecting on the generalizability concept. Indeed, method should well be considered a design facet for purposes of establishing trustworthiness. Research that is valid for practical applications is ideally grounded in a coherent conceptual framework and is multifaceted both with regard to designs and methods. Our investigations have aimed toward this ideal through multiple, overlapping, and intertwined approaches more than "grand experiments."

Pragmatics in Applying Research to Practice

The third theme centers around the possibilities and problems of applying "what we know," realizing that knowledge is always imperfect. Given a research base of the highest quality, engineering is required to fit the results to new and different settings. Primary among the challenges to this task in education is the disconnect between the worlds of research and practice, which is arguably greater in this field than other professional domains (Slavin, 2002).

Our experiences suggest that at least two phenomena are at work here. First are the ongoing debates about the roles and responsibilities of educational researchers. The camps include (a) those who see themselves as pure

scientists whose job is to search for “true” knowledge; (b) others who take a descriptive approach, documenting the successes and writhings of the clients; and (c) social engineers whose aims are pragmatic. Each perspective possesses functionality, but substantial energy goes into arguments about the relative merits of searching for the truth (cf. Anderson, 2002; Metz & Page, 2002, for a significant exchange). Whatever the merits of the various arguments, for practitioners, the result is a cacophony that discourages trust in any general pronouncements. From the beginning, then, efforts at applications appear fragmented on one side and skeptical on the other.

The second phenomenon arises from the “decoupling” typical of the school systems in the United States. In general, efforts to impose standard operating procedures at any level resemble herding cats. Once the classroom door closes, the teacher steers the ship of students with considerable autonomy. The managerial tools and incentives available to principals are limited except for programs that take shape as rigid scripts, with patrols enforcing the rules—a tactic that raises other problems. The more remote a mandate from the classroom, the less the response resembles the intended outcome. The national infatuation with Standards illustrates the point. In California, challenging academic standards were developed several years ago for the basic curriculum domains. At the same time, a standardized multiple-choice test with no basis in the standards was imposed to establish accountability. Today, when California educators emphasize teaching to the standards, they mean preparing students for multiple-choice items associated with one or more standards objectives.

The bottom line is that the educational system lacks stable points of leverage for program development, evaluation, and application. Under the best of circumstances, a district superintendent or a board of education, convinced of the merits of a particular program for whatever reason, directs implementation in one or more schools. Earnest discussion with principals reassures the program developer or researcher that the project is on track, teachers buy in, and all seems well. Even under these conditions, the activities and outcomes at the school, classroom, and student level, if anyone bothers to observe, can vary substantially from the original conception: sometimes at the surface level but more often in the deep structure. That is, all participants may do what they think they are supposed to do, but without understanding of the fundamental concepts, all of the pieces may be there, but they lack the skeletal structure.

For researchers, this state of affairs is a real problem in the sense of unwanted complications; to be sure, scholars interested in organizational issues make a living from such situations. Our experience suggests that this latter perspective may offer ways to address the matter. That is, researchers who decide to move toward application need to also study the implementation process. They can either develop expertise in the relevant domains or explore collaboration with other disciplines, a strategy posing challenges of

its own. At one time, the federal laboratory and center system (Vinoskis, 2001) seemed poised to support this strategy, but legislative support for the system has been erratic at best.

FOUR QUESTIONS

In this section of the chapter, we attempt to address four questions incorporating illustrations from the three projects, weaving in the themes along the way. The questions suppose that as researchers, (a) we aim to translate research into practice and (b) we actually know how to accomplish this task. Since the early 1980s, our team has indeed been interested in translation and believe that we have learned important lessons during these efforts. Whether we are actually prepared for this task is open to question, of course, especially when it comes to the task of scaling up. We were not prepared for this task by graduate work; and for academic researchers, engagement in practice can be hazardous for their publication records given the difficulty of such tasks and the time and energy needed to move from A to B.

The projects we discuss in this chapter were initially based in more than two decades of fundamental investigations grounded in the social-cognitive revolution of the 1960s and supported by a positive funding environment. In the early 1980s, the time seemed right for extending basic research to more complex settings, beyond what Glaser (1978) once characterized as “hot house” situations. The aim was not to expand a program, a popular idea in the 1960s (e.g., the First-Grade Reading studies; Bond & Dykstra, 1967), but to learn. How did the fundamental concepts and techniques play out over time and space, over contexts and groups? READ-Plus emerged as a research platform, an opportunity to extend the scientific agenda. For this reason, the task clearly fell within our province as researchers.

More recently, federal agencies have advanced the notion that researchers should act entrepreneurially to scale up their ideas and activities. Such a request is quite intriguing and can entice educational researchers to become advocates for their ideas and efforts. Without claiming to have completely avoided such temptations, we now approach the four questions from the stance that our aim is to advance a research agenda, which means grappling with the thematic issues raised earlier.

Question 1: Problems in Translating Theory and Research Into Practice

We first unpack three essential elements in this question: theory, research, and practice. Theory can be viewed from three stances. First is the academic perspective in which conceptual analysis provides an essential foundation for virtually all scholarship. A second emphasizes the underpinnings that

connect research with practice. The third appears in complaints from practitioners that “All ‘they’ talk about is theory; just tell us what to do.”

Research encompasses several constituents: empirical findings, methodologies, interpretations, and application. Schools must select research-based reading programs. A publisher may conduct rigorous evaluations to demonstrate positive learning outcomes, a relatively rare event. Other parties may employ the program in pursuit of other questions, and supporters can certainly use the data to advantage. Districts, schools, and teachers may also conduct investigations through local review of programs, pilot testing of materials, and occasional long-term evaluations; our experience is that such studies are discounted in favor of external, decontextualized reports. Local research may appear in anecdotal form in practitioner outlets, but less often in peer-reviewed journals.

Finally, practice takes different shapes. It can refer to a developer’s success in demonstrating the successful application of laboratory findings in school settings over an extended time. Given adequate resources and administrative support, most programs can manage this level of translation with positive outcomes, especially when the indicators are defined by the program and results are short-run (Goldberg, 2003). A different translation occurs when practitioners adopt ideas and techniques, using as part of their argument their review of research evidence. Players in this scenario can include teachers, school administrators, or policymakers; examples range from whole-language activists to phonics advocates.

Four learnings have emerged from our work in this complex arena. We do not characterize the following items as problems in the sense of unwanted difficulties hindering our efforts but rather as opportunities for extending the scope of our research.

Instruction as Learning Rather Than Activity. The first point reflects the emphasis in our work on concepts rather than activities as the basis for improving literacy instruction. Activity is captured in the IRE sequence; the teacher manages a time-filling routine that proceeds regardless of the outcome. A middle school incident illustrates the activity concept. Students have written geometry proofs on the board, and the teacher moves along evaluating each sketch. A student inquires, “Every ‘given’ has to be there, right? I think that C needs to be there as the center of the circle.” The teacher responds, “That’s right.” He modifies the proof and moves on with the activity. A conceptual response would have stopped the process: Why is C important? What happens to proof validity? And why did the student ask the question? However, the clock was running.

The aim in each READ-Plus program is a shared understanding by teachers (and administrators) of fundamental ideas about language, literacy, and learning, which then provide the basis for adapting curriculum sequences

and applying instructional strategies. In countries where education is viewed as a profession, the idea that knowledge is the basis for action provides a more friendly environment for this approach. To be sure, when the cognitive leap takes place for an individual practitioner, the result is a permanent change. At the outset, however, this mismatch between prevailing practice and the program foundations can lead to puzzlement and tension. Workshop evaluations complain frequently about “too much theory” and requests to “tell me what to do.”

An incident in the early READ-Plus workshop illustrates the point. At the end of each session, teachers were asked to react to the morning’s activities. As noted earlier, our team was new to the workshop business, and the focus on learning research probably seemed strange to the teachers. At the end of the second session, one teacher was bold enough to say, with some passion, “I don’t understand most of what you’re talking about—memory and meta-something and ‘chunking.’ And I don’t see how any of this stuff is going to help me teach reading this fall! But I’m staying, because we made a commitment—just wanted to let you know....” Several heads nodded agreement; the event was critical in surfacing both reactions and emotions. The opening discussion the following day continued with questions about bridging ideas and practices. Our team possessed expertise of one sort, but we clearly were in need of the “wisdom of practice.” The workshop was transformed by this event, which opened the opportunity to talk about literacy not as a basic skill but as a tool kit for problem solving and communication, the basis for the bridge that supported implementation during the following school year.

A central issue here is the assessment of student learning. For researchers, assessments are an integral part of program design; for administrators, assessments are tests; and for teachers, assessments are less clear-cut. In the translation to practice, two challenges emerge from our experience. First is the reliance in practice on surface-level indicators. The lesson—an event lasting an hour or less—is the centerpiece. If students complete the lesson activity, then learning has happened. Understanding and transfer seldom enter the picture. The second centers around the conception of individual differences. A common assumption in practice is that students vary in their potential, which leads to organization of the classroom into instructional groups, each associated with particular pedagogical practices and expectations. READ-Plus programs emphasize the commonality in human potential, more attuned to Brunerian/Bloomian notions that all students can attain significant achievements given appropriate instructional settings, time, and resources (Bloom, 1981; Bruner, 1966).

We assume that the same principles apply for adults, including teachers. Implementation of educational concepts and practices depends on learning outcomes. As we argue in our response to Question 3, success depends on time to internalize the concepts—time for practice with feedback in so-

cial settings that allow reflective interactions. That leads to the second topic in this section.

Sustaining Engagement. A significant challenge, echoed by many colleagues, is the task of sustained engagement with and by districts and schools. Both prepositions are important. *With* refers to leadership changes needed to support agreements. *By* recognizes the tumultuous bureaucratic seas in which administrators operate. Educational reforms entail 3 to 5 years for implementation, evaluation, and institutionalization. Changes take different shapes at different levels of the hierarchy. Some READ-Plus concepts can be applied with positive effects at the classroom level in a matter of days or weeks. Expertise and automaticity take more time, requiring school-level support and collaboration. School-wide implementation typically takes years of stable leadership; a new principal with a new agenda can change the context overnight. District-level policies and actions have emerged in our experience as perhaps the most significant factor for sustaining program development and implementation—and the integrity of associated research endeavors. The cabinet—associate superintendents and coordinators—plays a critical role. They determine principal placements, allocation of resources, and establishment of priorities. Superintendents and boards may come and go, but cabinets remain. Sustained engagement is also complicated by the inefficiency of research teams including our own efforts (Calfee & Patrick, 1995). Schools have short attention spans for understandable reasons. Universities do not prepare researchers for the pragmatics of program implementation, and our team has learned on the job.

The Nature of Evidence. The third problem centers around a cultural divide in views of the meaning and value of evidence. On one hand, data about student learning have limited impact on how schools operate, even in this age of accountability. Yearly reports of standardized test performance capture attention for a fleeting moment. Reactions fall into two categories. First is the “in your face” trend: Did scores go up or down? Gains are attributed to recent actions; declines call for something new. Second is relative rank: Schools with low scores are populated by students from poor families, a situation beyond the control of schools. The evidentiary base on which schools operate is sparse, most often externally mandated, multiple-choice tests and spring tests that arrive too late for action in the fall.

Other views of evidence build on a balance between externally mandated and internally generated sources of information, recognizing the several contrasts that distinguish these sources (Cole, 1988): authenticity of assessments, frequency of information gathering, availability of results, and linkage to curriculum, among others. Creation of longitudinal databases that portray trends is largely ignored in both external and internal assessment

programs. Education is about learning, and one might think that as a matter of course, educators would look at students' progress through the years. Testing companies typically deliver the results in yearly chunks, however, leaving it to districts, schools, and teachers to link achievements across the years.

A final point about evidentiary matters is the scarcity of local evidence about significant factors influencing student learning. Teachers matter, but they are seldom identified; the value-added method offers an approach along these lines but is by no means a commonplace. Innovative programs are introduced to make a change, but systematic local evaluation of these programs is relatively rare, nor is local evidence necessarily used in making local decisions. A collaborative research-based innovation is conducted between a researcher and a local district. Results are positive, long-lasting, and appear to transfer to a broad range of academic outcomes. The researcher is enthused, and the results are published. However, whether the innovation is retained is likely to depend not on the evidence but rather the vagaries of the local context. Politics and procedures are more likely to carry the day than scientific evidence. Researchers are tolerated but generally play a negligible role in decisions.

Resources for Research and Evaluation. A fourth barrier springs from the limited resources available for educational research and evaluation. The low importance placed on educational research is reflected in funding levels both locally and externally. Federal investments are a fraction of a percent of the educational enterprise. States and districts conduct mandated assessments but rarely go beyond required reports. Outside research activities must pay their own way, often including released time for teacher attendance at workshops. As Slavin (2002) points out, districts increasingly request bonuses for participation in experimental studies. The costs for preliminary pilot studies in limited and voluntary settings may be relatively inexpensive, but the researcher who aims to translate using large-scale studies will confront serious financial challenges. We might imagine a coordinated strategy involving researchers, districts, along with state and federal governments in which the design and collection of data became a collaborative enterprise. Some efforts can be found along these lines (the First-Grade Reading Studies attempted this approach; Bond & Dykstra, 1967), but the starting and ending points typically rest on data that are already available from bureaucratic mandates, which may not inform scientific questions.

Question 2: Ways to Overcome the Problems

The challenges we described previously are genuinely daunting, and one can understand why educational researchers might choose to conduct their basic work and let someone else worry about application. After all, the acad-

emy does not necessarily recognize the value of practical application in the field of education and may even disparage such efforts. The key may be to transform application activities into genuine research. In the following, we discuss two strategies that we think hold promise from this perspective.

Embedding and Scaling. The first strategy centers around the idea of embedding theory and research in practice rather than conducting research as an external and largely decontextualized activity to be passed on to implementers and practitioners at the end of the project. In proposing this idea, we are not suggesting that researcher and practitioner become one and the same. Backgrounds and roles differ significantly, and appropriately so. The design experiment concept offers a useful approach for accomplishing this goal (Collins, 1992; Calfee, Norman, Trainin, & Wilson, 2001). This term has assumed several shapes in recent years; the features we emphasize are (a) clear conceptualization of the educational problem and an explicit design for exploring the problem, (b) professional collaboration with practitioner partners in the operationalization and implementation of the investigation, and (c) ongoing reviews of activities and outcomes and continuous adaptation of the initial design based on the reviews.

One result from this approach has been our discovery of the value of both scaling up and scaling down. An early experience from scaling up is described in Calfee and Patrick (1995) when READ-Plus attempted in the early 1980s to move from the original middle-class site in San Jose to a collection of downtown schools serving large proportions of students from poor homes for many of whom English was a second language. The issue was not whether the program might benefit these students, but how it could survive in the chaotic organizational contexts typical of such schools. The effort came at the behest of the district administration, but that did not help with on-site implementation. The bottom line is that, although this effort at scaling up was not successful, the team learned a great deal about the influence of organizational context.

The importance of scaling down is a more recent learning. Once a particular activity has been implemented across a significant number of schools and districts, it often makes sense for the researcher to step back for a thorough review in a limited number of settings to better understand process and product, success and failure, and possibilities and limitations. In offering this advice, we assume that the researcher retains his or her role as researcher, which may conflict with entrepreneurial pressures. The aim is not program refinement but a deeper understanding of the concepts.

For instance, in Year 3 of the WordWork project, the research team decided to downscale, working with six teachers who volunteered for intensive exploration of the program. Teachers and researchers met for a full day each month to discuss “nuggets and lumps” encountered during imple-

mentation of WordWork during the previous weeks. Teachers provided data at each meeting on the progress of focal students along with details of their experience with program concepts and procedures. The mantra guiding the meetings was straightforward: By year's end, every student would meet district reading standards. Each meeting included a critical question springing from weekly classroom observations/interactions by the research team. Because the teachers had a year or more of experience with the team, they understood the place of critical review, and each discussion quickly took a life of its own, as teachers contributed their own questions and reflections. They developed remarkable adeptness at playing with ideas and offering suggestions but also challenging long-held beliefs. For example, the procedure for merging articulation with word building ran counter to teachers' previous experience with phoneme awareness ("Which word begins with the /b/ sound, *pat* or *bat*?") and phonics patterns ("Today we will work with '-at' words"). The engineering of popping/hissy sounds and glue letters was a continuing challenge for the research team. As the year moved ahead, enhancements of this program component came not from theory but from reflection on practical experiences.

The teacher meetings also covered more general matters that proved critical for program implementation. Especially significant were classroom management and grouping arrangements. The more experienced teachers had relied on ability grouping for decades; the idea of heterogeneous groups flew in the face of established beliefs and practices. The metaphonic concept at the core of WordWork requires student interactions that are difficult to sustain when students at the early stages of learning are placed together; none of the students can contribute to the discussion. A significant breakthrough came when one reluctant participant finally agreed to experiment with a mixed-ability group. Her subsequent report on the outcome reflected a genuine aha that altered her thinking and led to observable changes in her classroom practice; it also affected the entire group. The scientific import of this anecdote (by no means the only instance from this category) was not to inform the program design but to identify critical elements affecting program implementation. Establishing the effectiveness of a treatment for a traumatic disease like Alzheimer's is an important research goal. Equally important for doctors and patients is uncovering the pragmatics for implementing the treatment, which is likely to require scaled-down research—but research nonetheless.

As the year progressed, the research team documented increased depth of understanding as teachers discussed what worked and what didn't within the WordWork framework. The exchanges were grounded in shared terminology and perspective and in a focus on evidence-based student learning. Indeed, students made substantially greater progress on district indicators than similar students in previous years. By year's end, the teachers were able to demonstrate that virtually all their students had indeed met district stan-

dards. As a research activity, the exercise was an important demonstration, but it also generated significant refinements and understandings in the basic framework. These outcomes came about through the scaled-down strategy.

The RWS project also downscaled during its 3rd year. The five teacher participants during this phase exemplified deepening of understanding. Particularly interesting was a ripple effect that occurred during the year. As teachers gained greater confidence with the theoretical concepts, they were more effective at interacting with one another and sharing ideas with peers and the principal. The result was greater school-wide interest in RWS strategies along with a heightened awareness by district teachers and the administration about the relation of critical literacy to support for content area curricula. As it turned out, California's budget crisis took center stage at this point, ending the project.

Multilevel Connections. The second strategy calls for engaging practitioners across organizational levels as participants in the design experiment process. Embedding at the level of the classroom teacher, which we illustrated previously, is critically important. However, to understand contextual interactions requires going beyond self-selected volunteers to wrestling with the hierarchy. Most difficult in this regard is the district, especially in a time of intense accountability to external mandates; next is the school, given that leadership is subject to continual change; easiest is the teacher, assuming that access can be established.

A fundamental tension in the READ-Plus model, characteristic of most sociocognitive models, centers around the concept of professional teaching communities, which flies in the face of the hierarchical organization of American schools. At one point in the evolution of the model, we revived the concept of the "inquiring school" (Schaefer, 1967) in which we proposed that full-fledged literacy might provide the basis for genuine and equitable interactions among the various levels of the hierarchy. Both teachers and administrators often comment that "they didn't have time to think about what they are doing." Equally important, from our perspective, different levels of the system employ different concepts and technical language to describe their work. For the district, the curriculum consists of the textbook adoptions; for the teacher, it means moving through the daily lessons, picking up pieces along the way. For a district, assessment refers to standardized tests; for the teacher, it is a first grader's ability to make his or her way through a sentence during round-robin reading. The elementary school principal faces the challenge of bridging these gaps.

A practical illustration of the mismatch is found in the teacher's role in selecting curriculum programs. Contractual agreements allow teachers a voice in textbook adoption. Anecdotes at school board meetings can have significant impact. However, the discourse is informal and experiential,

with little reference to scientific evidence, published or local. The result is that the district cabinet often manages the actual decision.

In the WordWork study, for instance, 3rd-year teachers presented their work, including student outcomes, to a meeting of district and school administrators, which led to a public commitment by the district to a 3-year implementation of the program. On reflection, it is not clear that the presentation and the commitment were on the same page conceptually, but a year later, the district decided to implement a new basal adoption, in response to a state mandate, ending the earlier commitment. As researchers, we are still not sure how to interpret these findings. On one hand, for teachers and students, the achievement outcomes were impressive. However, as a study of scaling up, what would be the best way to move from these results to formulating the next study in the sequence in which the problem also encompasses the investigation of scaling up? Large-scale field experiments typically overlook local variations like those we just described. The assumption is that randomness will average out variations. We argue that understanding the nature of such variations is an important scientific goal in its own right.

Question 3: Success of Strategies

Success can be defined in various ways: (a) How many classrooms are presently using research-based Program X?, (b) Given that test scores have improved, who is benefiting—success for whom?, (c) What is the staying power of Program X at the classroom, school, and district levels?, and (d) What has been learned from the various programmatic activities that contributes to basic and practical understanding? Question 3 may also intend to direct attention to the strategies identified in the previous section: (a) embedding research throughout the process, and (b) connecting research activities across organizational levels.

A generalization stance means attending to a range of facets—people and organizations, learning outcomes, and time. Focusing for the moment on students, systematic, long-term follow-up encompassing a range of educational outcomes is difficult, especially in the absence of district engagement and especially for children from low-income homes. These students, the center of much of our effort, are prone to mobility, often moving within a district but sometimes within and across states; mobility can have a quite different meaning for English language learners. Nonetheless, we have persisted in pursuing students for 3 to 5 years using both standardized and project-based outcomes. Tracking such outcomes is challenging and does not fit neatly into the typical field-based experiment paradigm, which tends to emphasize quantity more than quality. Nonetheless, determining the success of a program—both the outcomes for students and the quality of the research endeavor—entails serious efforts to obtain such data.

A similar remark holds for adults. The National Reading Panel report (Langenberg, 1999) found little acceptable research on teachers and did not even mention administrators. We have pursued the hypothesis that substantial and long-term impact of any literacy program will depend on how it affects teachers—over the years, we have added administrators to this facet. Our experiences have documented a variety of remarkable multiyear transformations at the teacher and school level. As noted previously, all the school-level programs were eventually undone by district actions. Many teachers sustained the program concepts at classroom level, but teacher collaboration disappeared. Our conclusion is that it is possible to produce enduring changes in teacher knowledge and classroom practice that transcend particular programs, assignments, and contexts. However, moving from the classroom to the school level is essential, in our judgment, to provide students with the cross-grade consistency that is helpful for all children but essential for those most in need of academic support.

In the following, we present three snapshots that illustrate different perspectives for defining and evaluating the notion of success, bringing into play along the way the four points raised in chapter 1 (this volume).

A Principal With Vision. In *Teach Our Children Well*, Calfee and Patrick (1995) describe a 5-year READ-Plus project at a school serving a distressed neighborhood near Stanford University. Whether gauged by standardized measures of achievement, student and parent enthusiasm, reactions by the teachers, or attention to the program by the district, this project reflects the kind of bright-line event featured in practitioner journals like *Educational Leadership*. It was distinctive in several ways: (a) the largely unfunded activity was a collaborative endeavor, fostered by strong support from a remarkable principal who created conditions that led to the engagement of the entire school faculty; (b) the project continued to thrive well beyond any level of assistance from the original READ-Plus team; (c) the benefits for students continued when they moved on to middle school, and (d) the benefits for individual teachers continued when the principal was abruptly shifted to a different assignment.

The project demonstrated the effectiveness of a professional development program to dramatically enhance the performance of students from impoverished backgrounds. Over time, student performance as measured by standardized measures actually surpassed other district schools from more privileged backgrounds and not by changing the school's demographics. The children were still "from the 'hood.'" Equally important, from our perspective, was the sense of student efficacy evident in classroom activities and in discussions with individual students. Such reports are not unique to our program to be sure, but they can be evanescent and fleeting.

Eventually the project failed when the principal was reassigned. With the arrival of a new principal, connections were broken, teachers closed their class-

room doors, and opportunities for collaboration came to an end. Not exactly a new story, and what did we as researchers learn? From this instance in the mid-1980s, the answer was the initiation of a research question that we have yet to fully formulate but that underlies subsequent investigations under the WordWork and RWS banners: In what ways can professional development activities incorporate elements that support continuing school-level collaboration among teachers under prevailing conditions including changes in school and district contexts?

We continue to wrestle with how to conceptualize this question. Our initial ideas built on Schaefer's (1967) concept of *The Inquiring School: Literacy for the Year 2000* (Calfée, 1992). However, rather than basing inquiry on social science methods, our notion was to use literacy as the foundation; teachers with a shared vision of literacy as the capacity to solve problems and communicate already possessed the tools for school-wide collaboration! Perhaps so, but political savvy is needed as well as knowledge and skill—in the United States, unlike countries where the school is led by a head teacher, teachers do not perceive themselves as possessing professional authority. This chapter is not the place to explore the matter, but the challenge is how to reconstitute professional status in a conservative system that for decades has relied on the factory model. As things now stand, reform depends on remarkable leadership outside the classroom.

WordWork as a Targeted Activity. Unlike READ-Plus, which sought to transform all aspects of literacy instruction for the entire elementary school, WordWork focused on a delimited practice within a segment of the curriculum and for a subset of teachers. Development of the project was driven by practical as well as scientific considerations: phonics instruction was becoming a state emphasis; research support was available in this arena; and primary teachers told us that they needed more assistance in applying the READ-Plus concepts in the decoding-spelling area, an understandable request given the limited background available from current preparation programs (Calfée & Scott-Hendrick, 2004). Conceptually, our goal was to explore the metaphonic principle, the idea that acquisition of English orthography by young students would be enhanced by an emphasis on understanding rather than reliance on rote practice.

Accordingly, we proceeded to amplify the decoding-spelling segment of READ-Plus as a research platform. The focus was on classroom practice—curriculum, instruction, and assessment. The clients were primary grade teachers, lessening the pressures attendant to school-wide change. The curriculum was limited to phonics; to be sure, in the primary grades for schools serving high proportions of students from disadvantaged backgrounds, phonics can dominate the school day. The conceptualization centered around effectiveness and efficiency—reaching all students while

ensuring adequate time for parallel growth in vocabulary and comprehension. School-wide change was not part of the agenda, although the program emphasized collaboration among the primary-grade team.

At one level, the experiment was a success. Teachers who participated in the WordWork project deepened their understanding of fundamental concepts and became more effectual in applying their evolving knowledge and beliefs to practice. Significant changes were observed in classroom discourse during WordWork lessons, a marked departure from teacher-dominated talk (e.g., the IRE discourse structure) toward greater opportunities for students to interact with one another as well as the teacher in exploring orthographic concepts. In discussing spelling-sound relations, students performed as “little linguists” as they explained the functions of consonants and vowels.

These transformations were scaffolded by significant professional development activities. Teachers observed one another’s instruction and reflected on transcriptions of classroom discourse talk from WordWork lessons. These activities led them to insights about discourse patterns and the effect of teacher questions on student engagement and understanding. Traditional teacher-student roles—what gets said by whom—shifted when students had time to explore spelling patterns and reflect on their findings. In this context, time had two meanings: A big objective like “the short-A glue letter” would be the phonics focus for 2 weeks rather than only a single lesson, and during this time frame would also be part of the background agenda throughout the entire day, not just the 15 to 20 min of the official phonics lesson.

Equally important, transformations in teachers’ beliefs about teaching and learning led to expectations for elaborated student responses including explanations of their thinking. For example, in one of the monthly seminars, a teacher explained that she could “see the students thinking—WordWork teaches the kids thought processes, just like it is teaching the teachers to think.” Another teacher added that he now required students to “do the work, to manipulate letters and sounds and reason out why they go together like they do.” He remarked that too often the teachers do the students’ work.

These reflections are important in revealing the shift from activity to learning observed in most WordWork classrooms. This shift seemed to persist over time, well beyond the project’s tenure. The most obvious indicator was in “teacher talk”; teachers took less time and shaped their questions to evoke broader cognitive responses. When asked to explain what was going on, their intended focus was on finding out how students were thinking about a particular objective.

At several school sites, primary-grade teachers moved toward the creation of a Primary Team, joining kindergarten through second grade as a developmental group, a model employed in other countries but unusual in the United States. WordWork is designed as a spiral curriculum in which students move through the same orthographic concepts during the first 3

years of schooling. The spiral makes sense for young children and also promotes teacher collaboration. This model goes against the grain of current practice in which teacher manuals differentiate among grades; kindergarten is for readiness, first grade is the time for learning to read, and by second grade, the emphasis shifts to spelling.

Over several replications at different school sites, several WordWork teachers (their label) became recognized for their skill and knowledge in the area of early reading instruction. They presented at state and national conventions, served as district resources, and were notable for their professionalism in personal conversations. Classroom visitors were impressed, but as school and district change agents, their impact was limited. Success in this mix of outcomes can take many forms, but for the research agenda, scaling down was critical for investigating the process of teacher development.

RWS—Locating the Target. In the RWS project, the primary elements taken from READ-Plus came from the areas of vocabulary and expository text during the later elementary and middle school grades, with the intention of integrating language and literacy learning in the area of science. For readers familiar with the U.S. school system, the challenge of locating the target in this mix is apparent. Teachers in these settings tend to compartmentalize instruction, driven by instructional schedules akin to those in high schools. Even in the early elementary years, the concept of integrated, interdisciplinary instruction can be subverted by the textbook system and the daily schedule. “How are we going to have time to do ‘this’ too?” was often the first response from teachers asked to consider the project. A strategy designed to integrate elements on the daily menu was viewed as one more thing to do.

Once over this hurdle and with some consistent administrative support, an appreciation that integrating reading and writing in the content areas could actually save instructional time eventually became apparent to many teachers, sometimes by the end of the first curriculum unit. For most participating teachers, concepts of text analysis, rhetorical structure, graphic organizers, vocabulary development, and expository writing were genuinely new. The initial workshop sessions on theory were, as usual, a challenge. The rubber hit the road during full-day meetings held prior to each unit to review and discuss specific components. During these sessions, teachers worked through lesson plans, assessment materials, curriculum mandates (how to cover the textbook), and ancillary resources from the library and Internet.

In later phases, RWS scaled down the formal agenda for these sessions, encouraging teachers to become more active participants in curricular adaptations. Although teacher collaboration was an important facet of the program design, the reality was that interactions were largely limited to project meetings. Unlike WordWork in which primary grade teachers were eager to compare notes, the RWS experience suggests that upper elemen-

tary teachers are less likely to seek out occasions for collaboration—perhaps reflecting the enormous burden they bear as teachers of everything.

Administrators at the school and district levels found the project interesting, but this constellation of curriculum topics was not high on their priorities. In California, reading and math are the centerpiece for state testing and school accountability. Writing is tested at selected grades, and science is part of the mix, but neither receive much weight in the current school accountability index. Context matters.

In summary, instructional strategies falling under the general READ-Plus umbrella have proven effective for students, especially those who arrive at school lacking fluency in academic language. Professional development strategies, where these can be implemented for a sustained amount of time, appear to lead to fundamental changes for teachers. However, context is a critical factor. Efforts to deal with organizational issues at the school and district level—through sustained engagement, attention to evidence of instructional practice and student learning, and efforts to promote local support for research and evaluation—have been far less effective in our experience. Research on large-scale implementation may require small-scale investigations.

Question 4: Lessons for the Field

So, back to the title: What have we learned from our experiences about the relation/translation of educational research into the realm of practice? Our answer must necessarily be contextualized within the contemporary arenas of policy and practice. We couch the lessons against the three themes presented earlier.

Three Themes Redux. If one is to accept current congressional and administrative promulgations, educational research is actually quite simple: follow legislated methodological mandates. Our experience suggests otherwise, hence the theme of research as complex learning. We think that a careful reading of the record of educational research during the past half century supports this conclusion. Research is clearly complex, whether in physics, biology, medicine, engineering, or education. The challenge is greater in those fields, such as the latter three, that emphasize practical application of fundamental findings.

At the core of any research effort are three elements: conceptualization, methodology, and the empirical base. In education, the conceptual base has substantial potential but is scattershot to say the least. Education is a political endeavor, and so the emphases can vary widely; environment and ecology pose similar challenges. Should education be conceived as the acquisition of prescribed skill and knowledge or the development of the educated individual capable of contributing to a democratic society?

Methodology follows on conceptualization. Demonstrating the statistical superiority of particular programs for the attainment of specific outcomes calls for one set of methodologies; understanding the underlying processes of learning calls for a different set of methodologies. Joining these two endeavors, which seems desirable, calls for combined methodologies.

The evidentiary base for education has for several decades been catalogued in the Educational Resources Information Center, with relatively permissive criteria for entering documents. More recently, policymakers have determined that the empirical base in education—the database for researchers and decision makers—should be proscribed, limited to particular problems and methodologies. Our experience suggests that such proscriptions may not be in the best interests of researchers or practitioners or the public. Empirical resources presently far exceed our reach in most areas. The quality is mixed, to be sure. However, the lesson would seem fairly straightforward: More is probably known about improving practice and enhancing student outcomes than is presently realized in practice—conceptually, methodologically, and empirically.

A second theme focused on the issue of research that can be applied to practice. The body of research on how humans learn that has accumulated during the past several decades is impressive beyond belief (Bransford, Brown, & Cocking, 1999). Application stands forth as the challenge—what should be applied, for whom, and with what resources. Decisions about valued outcomes are central to approaching this matter. R. Tyler (personal communication, June, 1991) once commented that “education comprises the means by which a society passes on its values from generation to generation.” The starting assumption for our work rests on the notion that American education should promote the capacity of every individual to play a significant role in a democratic society, which entails independence, life-long learning, and consideration of others. Another assumption centers on scientific validity, the rigor of conception, design, methods, and interpretations of contributions to the research base, with particular attention to generalizability. The lesson we take away here relates in part to the engineering task we mention next, but also to the need for adequate support for applied and basic educational science.

The third theme centered around the practicalities of application—the engineering task. This theme accents the contrast between academic and practitioner environments in education. In other domains—engineering and medicine come to mind—the connections appear more workable; in education, the chasm seems greater. Federal policy can exacerbate the problem when it endorses particular problems and methods and directs practitioners to follow external mandates. Such policies finesse the engineering task.

Despite the plethora of meta-analytic reviews available in a variety of domains, including reading, the translation of findings into practical advice

seems rather limited. As an example, a recent review of the impact of direct-instruction programs in the primary grades suggested limited transfer to reading comprehension in the elementary grades (Kemper & MacIver, 2002; also Wenglinisky, 2003). It seems unlikely that most teachers and administrators will have read the review. More often, a program is presented to school districts as research based with a sprinkling of citations. Local research and evaluation of program effectiveness become critical at this point. More likely are the newspaper reports in which a district or school uses a jump in reading scores to justify a particular program. To be sure, local expertise is generally sparse, except in large districts. Collaboration with academic partners might help, but resources for research and evaluation are also sparse, and the participants live in different worlds. Except for occasional large-scale projects, the typical paradigm is the request by a district or school for a local campus to "find a grad student" who can conduct an evaluation, perhaps as a dissertation project. In the spring of 2003, the National Research Council called for a state-based strategic research partnership, but the half-billion-dollar cost seems unlikely in today's fiscal environment ("National Academy Calls," 2003).

The lesson from our experience is that exploring the effectiveness of ideas and practices—the local engineering—requires more than assistance from novices. It requires a system that combines sustained efforts from universities with the field of practice. At the University of California Riverside, agricultural research is a major activity, including studies of "turf grass," the playing field serving golfers in Palm Springs and thereabouts. Research on turf grass spans the spectrum from laboratory studies to field work, investigations that follow a new strain from genetic construction to putting surface in which the various players speak a common language and share common values (cf. Kohler, 2002, for a discussion of experimental field stations).

Our explorations of ideas and practices for improving the educational opportunities for youngsters at risk for school failure resonate with the experimental educational station concept—a system in which all parties establish common ground. Realizing such a visionary goal will call for resolution of three tasks raised at the outset—sustained engagement, agreement on valid evidence as a basis for action, and local involvement in systematic research and evaluation. These are important targets and warrant continued pursuit—by researchers interested in fundamental issues of learning science and by colleagues entrusted with implementation of effective policies and practices at the local level in schools and classrooms.

REFERENCES

- Anderson, G. L. (2002). Reflecting on research for doctoral students in education. *Educational Researcher*, 31(7), 22–25.
- Anderson, N. H. (2001). *Empirical direction in design and analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Bloom, B. S. (1981). *All our children learning*. New York: McGraw-Hill.
- Bond, G. L., & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly*, 2(4).
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Bruner, J. S. (1966). *Toward a theory of instruction*. Cambridge, MA: The Belknap Press of Harvard University.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.
- Calfee, R. C. (1992). The inquiring school: Literacy for the year 2000. In C. Collins & J. N. Mangieri (Eds.), *Teaching thinking: An agenda for the twenty-first century* (pp. 147–166). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Calfee, R. C. (1998). Phonics and phonemes: Learning to decode and spell in a literature-based program. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 315–340). Mahwah, NJ: Lawrence Erlbaum Associates.
- Calfee, R. C., & Associates. (1981). *THE BOOK for the reading teacher: Components of reading instruction. A generic manual for reading teachers*. Unpublished manuscript, Stanford University School of Education.
- Calfee, R. C., & Nelson-Barber, S. (1991). Diversity and constancy in human thinking: Critical literacy as amplifier of intellect and experience. In E. Hiebert (Ed.), *Literacy for a diverse society: Perspectives, programs, and policies* (pp. 44–57). New York: Teachers College Press.
- Calfee, R. C., & Norman, K. A. (1999). Psychological perspectives on the early reading wars: The case of phonological awareness. *Teachers College Record*, 100, 242–274.
- Calfee, R. C., Norman, K. A., Trainin, G., & Wilson, K. (2001). Conducting a design experiment for improving early literacy, or, what we learned in school last year. In C. Roller (Ed.), *Learning to teach reading: Setting the research agenda* (pp. 166–179). Newark, DE: International Reading Association.
- Calfee, R. C., & Patrick, C. (1995). *Teach our children well*. Stanford, CA: The Portable Stanford Series, Stanford Alumni Association.
- Calfee, R. C., & Scott-Hendrick, L. (2004). The teacher of beginning reading. *Contemporary Perspectives on Early Childhood Education*, 5, 87–117.
- Carr, E., & Ogle, D. (1987). K-W-L plus: A strategy for comprehension and summarization. *Journal of Reading*, 30, 626–631.
- Cazden, C. B. (1988). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.
- Chambliss, M. J., & Calfee, R. C. (2002). The design of empirical research. In J. Flood, J. M. Jensen, D. Lapp, & J. R. Squire (Eds.), *Handbook of research on teaching the English language arts* (2nd ed., pp. 152–170). Mahwah, NJ: Lawrence Erlbaum Associates.
- Chambliss, M. J., & Calfee, R. C. (1998). *Textbooks for learning: Nurturing children's minds*. Oxford, England: Blackwell.
- Cole, N. (1988). A realist's appraisal of the prospects for unifying instruction and assessment. In C. V. Bunderson (Ed.), *Assessment in the service of learning* (pp. 1–23). Princeton, NJ: Educational Testing Service.
- Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 37–53). New York: Springer-Verlag.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671–684.

- Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 30, 116–127.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Eisenhart, M., & Towne, L. (2003). Contestation and change in national policy on “scientifically based” research. *Educational Researcher*, 32(7), 31–38.
- Fisher, R. A. (1935). *The design of experiments*. Edinburgh, Scotland: Oliver & Boyd.
- Glaser, R. (1978). *Advances in instructional psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Goldberg, M. (2003). Everything works. *Phi Delta Kappan*, 85, 304–306.
- Heath, S. B. (1983). *Ways with words*. New York: Cambridge University Press.
- Kemper, E. A., & McIver, M. A. (Eds.). (2002). Direct instruction reading programs: Examining effectiveness for at-risk students in urban settings. *Journal of Education for Students Placed at Risk*, 7(2).
- Kincaid, H. (1996). *Philosophical foundations of the social sciences*. New York: Cambridge University Press.
- Kohler, R. E. (2002). *Landslopes and labscopes: Exploring the lab field border in biology*. Chicago: University of Chicago Press.
- Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.
- Langenberg, D. N. (1999). *Report of the National Reading Panel*. Washington, DC: NICHD.
- Lapp, D., Flood, J., & Hoffman, R. P. (1996). Using concept mapping as an effective strategy in content area instruction. In D. Lapp, J. Flood, & N. Farnan (Eds.), *Content area reading and learning* (2nd ed., pp. 291–306). Boston: Allyn & Bacon.
- Marshall, E. (2003). In critical condition. *Science*, 300, 1225–1226.
- Metz, M. H., & Page, R. N. (2002). The uses of practitioner research and status issues in educational research: Reply to Gary Anderson. *Educational Researcher*, 31(7), 26–27.
- Miller, R. G., & Calfee, R. C. (2004). Building a better reading/writing assessment: Bridging cognitive theory, instruction, and assessment. *English Leadership Quarterly*, 26, 6–13.
- National Academy calls for \$500 million strategic education research partnerships. (2003, April). *Science*, 300, 231.
- Natriello, G. (2004). Scientific research in education. *Teachers College Record*, <http://www.tcrecord.org>, ID 11165.
- Nelson, N. N., & Calfee, R. C. (Eds.) (1998). *The reading-writing connection: The yearbook of the National Society for the Study of Education*. Chicago: University of Chicago Press.
- Pearson, P. D. (in press). The reading wars: The politics of reading research and policy—1988 through 2003. In B. C. Johnson & W. L. Boyd (Eds.), *Politics of Education Yearbook, Special Issue on Educational Policy*.
- Phillips, D. C. (2000). *The expanded social scientist's bestiary: A guide to fabled threats to and defenses of naturalistic social science*. New York: Rowman and Littlefield Publishers.
- Prairie, V., & Hand, B. (1996). Writing for learning in secondary science: Rethinking practices. *Teacher and Teacher Education*, 12, 609–626.
- Reyna, V. F. (2004). The No Child Left Behind Act and scientific research. In J. S. Carlson & J. R. Levin (Eds.), *Scientifically based education research and federal funding agencies: The case of the No Child Left Behind Act* (pp. 3–25). Greenwich, Ct: Information Age Publishers.

- Robinson, V. M. K. (1998). Methodology and the research-practice gap. *Educational Researcher*, 27, 17-26.
- Sanders, L. (2003, March 16) Medicine's progress, one setback at a time. *New York Times Magazine*, pp. 29-31.
- Schaefer, R. J. (1967) *The school as a center of inquiry*. New York: Harper and Row.
- Shavelson, R. J., Phillips, D. C., Towne, L., & Feuer, M. J. (2003). On the science of education design studies. *Educational Researcher*, 32, 25-28.
- Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education*. Washington, DC: National Academy Press.
- Slavin, R. E. (2002). Evidence-based education policies: Transforming educational research and practice. *Educational Researcher*, 31(7), 15-21.
- Sperling, M., & Freedman, S. W. (2002). Research on writing. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 370-389). Washington: American Educational Research Association.
- Stanovich, K. E. (2003). Understanding the styles of science in the study of reading scientific-studies of reading, *Scientific Studies Research*, 7, 105-126.
- Tierney, R. J., & Shanahan, T. (1991). Research on the reading-writing relationship: Interactions, transactions, and outcomes. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research Vol. 2* (pp. 246-280). Mahwah, NJ: Lawrence Erlbaum Associates.
- Vinoskis, M. (2001). *Revitalizing federal education research and development*. Ann Arbor, MI: University of Michigan Press.
- Wenglinsky, H. (2003). Using large-scale research to gauge the impact of instructional practices on student reading comprehension: An exploratory study. *Educational Policy Analysis Archives*, 11(19), 2-15.