

Article 1

THE READING WARS¹

P. DAVID PEARSON

This article's fundamental argument is that reading instruction and reading research have been shaped by political forces desiring to privilege particular approaches to instruction or particular combinations of methodological and epistemological perspectives on research. The swings in both dominant pedagogies and dominant research paradigms are analyzed in terms of these determining forces. The article concludes by championing balance and compatibility across both instructional approaches and research methods in hopes of arresting the pendulum swings that have characterized the field for too many decades.

Keywords: reading; reading research; reading policy; history of reading; reading curriculum

Writing about the politics of reading some 15 years ago (Pearson, 1989), I wondered whether the *whole-language movement*, which was the centerpiece of the reading field's foray into constructivist pedagogy, was capable of maintaining the mantle of "conventional wisdom," a status that at that time, it was on the brink of achieving. I questioned its enduring leadership capacity because of the curricular, philosophical, and political ground on which it stood. Curricularly, I

expected that its guiding principles of authenticity (in texts, tasks, and tests) and curricular integration—both within the language arts (across reading, writing, speaking, and listening) and between the language arts and other curricular areas—would run afoul of the powerful publishing lobbies in the United States. Philosophically, it is built on epistemologies of interpretation rather than realism, rejecting the idea of an external reality that we will eventually find if we just look hard enough; as such, leaders of the whole

SOURCE: Pearson, P. D. (2004). The Reading Wars. *Educational Policy*, 18(1), 216–252. Reprinted by permission of Sage Publications, Inc.

language movement would have been thrilled to find provisional and situation-specific answers to burning policy questions such as, What is the best way to teach beginning reading? Those views would not sit well, I thought, in congressional or state legislative milieus or school board chambers, places where truth and simple answers to policy questions are serious goals. Politically, I predicted that its commitment to grassroots decision making—a commitment requiring that everything must be done to preserve as much power and prerogative for individual teachers (who must, in turn, offer genuine choices to individual students)—would doom it as a policy initiative. In an atmosphere in which accountability systems driven by externally mandated high-stakes tests lay just over the horizon, I wondered whether policy makers, or parents for that matter, would be willing to cede that level or prerogative to a profession that in terms of its capacity to deliver achievement, seemed to be asleep at the wheel. My overarching question was whether whole language could withstand the pressure of curricular leadership, with implicit responsibility for whatever trends in achievement ensued. My suspicion was that it was better situated as a guerilla-like movement that made occasional sorties into the policy world to snipe at those in curricular power.

In reflecting on those wonderments some 15 years later, it is clear that whole language, along with its close constructivist cousins—literature-based reading, process writing, and integrated language-arts instruction—did not experience a long tenure in the seat of curricular power, at least in the form in which it and its relations existed in the late 1980s and early 1990s. Whether the seeds of its demise were internal shortcomings, as I wondered in 1989, or external political forces of the sort that dominate the policy conversation today, or some combination of the two, is a question that I return to at the end of this article after reviewing the important developments in policy and practice that have shaped events and interpretations in the interim.

THE GOLDEN YEARS OF WHOLE-LANGUAGE INFLUENCE

Whole language did not suddenly emerge on the reading scene in the 1980s. Its roots (Y. Goodman, 1989) are in Deweyian-inspired, child-centered pedagogy and the integrated curriculum movements popular in England, Australia, and New Zealand (e.g., Holdaway, 1984). It also owes part of its heritage to earlier American movements, such as individualized reading (Veatch, 1959) and language experience (Stauffer, 1980). But it was the incredible shift in the scholarly paradigms that undergirded our views of reading acquisition that in my view, really laid the groundwork for its ascendancy (see Pearson & Stephens, 1993, for an account of these developments). The psycholinguistically oriented work of Roger Brown (1970), Frank Smith (1971), and Kenneth Goodman (1965, 1969) sent the message that reading was more a language than it was a perceptual process. The work in reading comprehension inspired by the cognitive revolution in psychology (see Anderson & Pearson, 1984) established meaning as the core, not the residual outcome, of reading. Advances in sociolinguistic theory in the 1980s (Bloome & Green, 1984; Heath, 1983) and critical literacy in the 1990s (Gee, 1989; Luke, 1995) established the understanding that all language and, hence, all literacy learning is grounded in the material motives of human interaction, with all of its social, political, and economic faces (however endearing or ugly they might be) intact.

When whole language emerged as a movement in the 1980s, it challenged the conventional wisdom of basals and questioned the unqualified support for early code emphases that had grown between 1967 and the early 1980s. One of the great ironies of whole language is that its ascendancy into curricular prominence is best documented by its influence on the one curricular tool it has most consistently and most vehemently opposed, the basal reader. Basals changed

dramatically in the early 1990s, largely, I am confident, in response to the groundswell of support within the teaching profession for whole language and its close curricular allies, literature-based reading and process writing.

Vocabulary control, already weakened during the 1970s in response to Chall's (1967) admonitions, was virtually abandoned in the early 1990s in deference to attempts to incorporate more literature, this time in unexpurgated form (i.e., without the practices of adaptation and excerpting that had characterized the basals of the 1970s and 1980s) into the Grade 1 program (Hoffman et al., 1995). Phonics, along with other skills, was backgrounded, and literature moved to center stage.

Basal programs appropriated or, as some whole-language advocates have argued, "basalized" the activities and tools of whole language. Thus, in the basals of the early 1990s, each unit might have a writing process component in which the rhetoric if not the reality of some version of process writing was presented to teachers and students. In the 1980s, comprehension questions, probably following a story line, might have sufficed for the guided reading section of the manual (the part that advises teachers on how to read and discuss the story), but in the 1990s, questions and tasks that supported deep probes into students' responses to literature became more prevalent. Another concession to literature-based reading was the creation and marketing of classroom libraries—boxed sets of books, usually thematically related to each unit, that teachers could use to extend their lessons and units "horizontally" and enrich children's literary opportunities.

Basals also repositioned their "integrated language arts" and "integrated curriculum" strands. Dating back even to the 1920s and 1930s, basals had provided at least a "token" section in which teachers were encouraged to extend the themes or skills of the basal story into related writing (e.g., rewriting stories), oral language (e.g., transforming a story into a play and dramatizing it),

or cross-curricular activities (e.g., conducting community surveys, tallying the results, and reporting them), but these forays were regarded as peripheral rather than core. In the basals of the early 1990s, as skills moved into the background, these integrated language-arts activities were featured more prominently as core lesson components.²

These changes can, I believe, be traced to the prominent position of whole language as a curricular force during this period (Pearson, 1992). Publishers of basals accomplished this feat of appropriation not by ridding their programs of the skills of previous eras but by subtle repositioning—foregrounding one component while backgrounding another, creating optional components or modules (e.g., an intensive phonics kit or a set of literature books) that could be added to give the program one or another spin. Unsurprisingly, this created bulkier teachers' manuals and more complex programs.

Acceptance of whole language was not universal. To the contrary, there was considerable resistance to whole language and literature-based reading throughout the country.³ In many places, whole language never really gained a foothold. In others, what was implemented in the name of whole language was not consistent with the philosophical and curricular principles of the movement; California, whole-language advocates would argue, is a case in point. Whole language got conflated with whole-class instruction and was interpreted to mean that all kids should get the same literature, even if teachers had to read it to them.⁴

Nor was there a single voice within the whole-language movement. Whole-language scholars and practitioners differed, and still differ, on a host of issues such as the role of skills, conventions, and strategies within a language-arts program. Some said, if we can just be patient, skills will emerge from meaningful communication activities; others spur things on by taking advantage of spontaneous opportunities for minilessons; still others were willing to

spur spontaneity a bit with minilessons and other transparently instructional routines.

Even so, it is fair to conclude that by the early 1990s, whole language had become the conventional wisdom, the standard against which all else was referenced. The rhetoric of professional articles belies this change. As late as the mid-1980s, articles were written with the presumption of a different conventional wisdom—a world filled with skills, contrived readers, and workbooks. By 1991–1992, they were written with the presumption that whole-language reforms, although not fully ensconced in America’s schools, were well on their way to implementation. The arguments in the 1990s were less about first principles of whole language and more about fine-tuning teaching repertoires. The meetings of the Whole Language Umbrella grew to be larger than most large state conventions and regional conferences of the International Reading Association. By 1995, whole language was no longer a collection of guerrilla sorties into the land of skills and basals that characterized it through the mid-1980s. It had become the conventional wisdom, in rhetoric if not in reality.

THE DEMISE OF WHOLE LANGUAGE

Toward century’s end, just when it appeared as if whole language, supported by its intellectual cousins (process writing, literature-based reading, and integrated curriculum), was about to assume the position of conventional wisdom for the field, the movement was challenged seriously, and the pendulum of the pedagogical debate began to swing back toward the skills end of the curriculum and instruction continuum. Several factors converged to make the challenge credible, among them (a) unintended curricular casualties of whole language; (b) questionable applications of whole language; (c) the growth of balanced literacy as a mediating force in the debate; (d) a paradigm shift in the ideology of

reading research; (e) increasing politicization of the reading research and policy agenda; (f) increasing pressure for educators of all stripes, especially reading educators, to produce measurable results; and (g) loss of the moral high ground. All of these forces, but especially those delineated above as *d* through *f*, came together in one place—the reading first component of the No Child Left Behind Act of 2001 (NCLB, 2002). By the time this article was submitted for publication, in mid-2003, reading first had assumed the role of conventional wisdom in reading instruction, albeit by mandate rather than groundswell, and only a few traces of whole language, which seemed so dominant only 7 years earlier, could be found in our schools and curricula. How did this remarkable political transformation occur? That is the subject of this article.

Unintended Curricular Consequences

In its ascendancy, whole language changed the face of reading instruction and in the process, left behind some curricular casualties, few of which were intended by those who supported whole language. Those, including many curricular moderates, who supported practices that were discarded in the rise of whole language had difficulty supporting the whole-language movement even though they might have been philosophically and curricularly sympathetic to many of its principles and practices (see Pearson, 1996). This lack of enthusiasm from curricular moderates meant that whole language failed to build a base of support that was broad enough to survive even modest curricular opposition, let alone the political onslaught that it would experience at century’s turn.

There were four casualties: skills instruction, strategy instruction, an emphasis on text structure, and reading in the content areas. Earlier, I suggested that one of the consequences of whole language was the relegation of skills to the “appendices” of instructional programs. In accepting whole language, we tacitly accepted the premise that skills are better caught in the act of reading and writing genuine texts for authentic

purposes than taught directly and explicitly by teachers. The argument is the same for phonics, grammar, text conventions, and structural elements. These entities may be worthy of learning, but they are unworthy of teaching. This position presents us with a serious conundrum as a profession. Admit, for the sake of argument, that the skills instruction of the 1970s and earlier, with decontextualized lessons and practice on “textoids” in workbook pages, deserved the criticism accorded to it by whole-language advocates (and scholars from other traditions). But a retreat from most skills instruction into a world of “authentic opportunity” did not provide a satisfactory answer for teachers and scholars who understood the positive impact that instruction can have. Many young readers do not “catch” the alphabetic principle by sheer immersion in print or by listening to others read aloud. For some it seems to require careful planning and hard work by dedicated teachers who are willing to balance systematic skills instruction with authentic texts and activities (see Hiebert & Taylor, 1994, for a description of many of the interventions designed to accomplish just this balanced goal).

Strategy instruction (intentional attempts to equip students with meta-cognitive routines for understanding text, monitoring comprehension, and fixing things up when they go awry) was another casualty. This loss was particularly difficult for scholars who spent the better part of the 1980s convincing basal publishers and textbook authors that the thoughtful teaching of flexible strategies for making and monitoring meaning was a viable alternative to ubiquitous skill instruction, where skills were taught as though they were only ever to be applied to workbook pages and end-of-unit tests. But the strategy lessons that filled basals in the mid- to late 1980s were virtually nonexistent in the basals of the early to mid-1990s. Although there is no inherent bias in whole language or literature-based reading against the learning and use of a whole range of cognitive strategies, there is, as with phonics and grammar, a serious question about whether direct, explicit instruction in how to use them will help. The advice is to let them emerge from

attempts to solve real reading problems and puzzles, the kind students meet in genuine encounters with authentic texts.

Structural emphasis was also suspect within whole language. This suspicion extended to formal grammars, story grammars, rhetorical structures, and genre features of texts. As with skills and strategies, whole-language reformers did not claim that students should not learn and develop control over these structural tools; they simply claimed that like skills, they are best inferred from reading and writing authentic texts in the process of making meaning. So, the advocates are comfortable in adopting Smith’s (1983) admonition to encourage kids to read like a writer (meaning to read the text with a kind of critical eye toward understanding the tools and tricks of the trade that the author uses to make points and achieve his or her effects on readers), but they would likely reject a systematic set of lessons designed to teach and assess children’s control of story grammar elements (such as plot, characterization, style, mood, or theme) or some system for dealing with basic patterns of the various genres of expository text. As with skills and strategies, many in the field sought a compromise alternative to both the formulaic approach of the early 1980s and the “discovery” approach of the new reforms—dealing with these structural elements as they emanate from stories that a group is currently reading can provide some guidance and useful tools for students and teachers.

Content area reading also suffered during the ascendancy of whole language and literature-based reading. Content area texts—expository texts in general, but especially textbook-like entries—were not privileged in a world of literature-based reading. There is a certain irony in this development, for it is competence with expository reading, not narrative reading, that most concerns educators and future employers. The cost here has been very dear. Concerned that students either cannot or will not read textbook assignments, most high school teachers have chosen either to read the text to students or even more likely, to tell students what they would have

encountered had they or could they have read it. Although understandable, this approach is ultimately counterproductive. There comes a time in the lives of students—either when they go to college or enter the world of work—when others expect them to read and understand informational texts on their own and in printed form rather than through oral or video transformation.⁵

Because whole language did not go out of its way to accommodate these structural- and content-focused curricular practices, those who were sympathetic with whole language but also champions of one or another approach were not available to help whole language respond to the criticism leveled at it in the late 1990s. Building allies across boundaries of curricular political divides was not, as it turned out, a strength of the movement.

Questionable Applications of Whole Language

One of the dilemmas faced by any curricular initiative is sustaining the integrity of the movement without imposing the very sorts of controls it is trying to eliminate. Whole language did not find a way to manage this dilemma, and it suffered as a consequence. Many schools, teachers, and institutions appropriated the whole-language label without honoring its fundamental principles of authenticity, integration, and empowerment. Basal-reader publishers made the most obvious and widespread appropriation, some even positioning their basal series as “whole-language” programs. The most egregious misapplication was the conflation of whole language with whole-class instruction. Nowhere was this conflation more extreme than in the implementation of the California literature framework. The logic that prevailed in many classrooms was that it was better to keep the entire class together, all experiencing the same texts, even if it meant that the teacher had to read the text to those children who lacked the skills to read it on their own. Implicit in this practice are two interesting assumptions: (a) that getting the content of the stories is the most important goal for reading

instruction, and (b) that the skills and processes needed to read independently will emerge somehow from this environment in which many students are pulled through texts that far exceed their grasp, given the sophistication of their current skills repertoire. Needless to say, whole language had enough on its hands dealing with its own assumptions and practices; these philosophical and curricular misapplications exposed the movement to a whole set of criticisms that derived from practices not of its own making.

A plausible explanation for the misapplication of whole language was its lack of an explicit plan for professional development. Given its grassroots political assumptions, it is not surprising that whole language gave teachers a wide berth for making curricular and instructional decisions. It assumed that teachers who are empowered, sincere, and serious about their work would be able to tailor programs and activities to the needs and interests of individual children. Such an approach makes sense only when teacher knowledge is widely and richly distributed in our profession. To offer these prerogatives in the face of narrow and shallow knowledge is to guarantee that misguided practices, even perversions of the very intent of the movement, will be widespread. The puzzle, of course, is where to begin the reform—by ensuring that the knowledge precedes the prerogative, or by ceding the prerogative to teachers as a way of leveraging their motivation for greater knowledge. Similar arguments have been made for the reform movements in mathematics (i.e., that the reforms got out ahead of the professional knowledge base); interestingly the reform movement in mathematics has experienced a fate similar to that of the whole-language movement (see Good & Braden, 2000; Schoenfeld, 2004.)

Balanced Literacy

Although it has reached its peak in the past 5 years, concern about extreme positions, be they extremely child centered (such as the more radical of whole-language approaches) or extremely curriculum centered (such as highly structured,

unswerving phonics programs), is not new. Voices from the middle, extolling balanced approaches or rationalizing the eclectic practices of teachers, began to be heard even in the earliest days of whole language's ascendancy.⁶ Scholars and teachers raised a number of concerns about the assumptions and practices of the whole-language movement. Most important, they expressed concern about the consequences of whole language outlined earlier in this article. They questioned the assumption that skills are best "caught" during the pursuit of authentic reading activity rather than "taught" directly and explicitly. They also questioned the insistence on authentic texts and the corollary ban on "instructional" texts written to permit the application of skills within the curriculum. They questioned the zeal and commitment of the movement *qua* movement, with its strong sense of insularity and exclusivity. Finally, they worried that the press toward the use of authentic literature and literature-based reading would eradicate, albeit unintentionally, what little progress had been made toward the use of informational texts and teaching reading in the content areas (Pearson, 1996).

Ironically, in the past few years, these voices from the middle have found themselves responding not to those who hold a radical whole-language position but to those who hold steadfastly to the phonics first position. Even so, the fact that those with centrist positions were not inclined to defend whole language when the political campaign against it began in the middle 1990s undoubtedly hastened the demise of whole language as the pretender to the title of conventional wisdom.

Changing Research Paradigms

Prior to the 1980s, qualitative research in any form had little visibility within the reading research community. Among the array of qualitative efforts, only miscue analysis⁷ and some early forays into sociolinguistic and anthropological accounts of literacy had achieved much in the way of archival status.⁸ However, all that changed in the 1980s and early 1990s. Qualitative

research more generally, along with more specific lines of inquiry taking a critical perspective on literacy as a social and pedagogical phenomenon, became more widely accepted as part of the mainstream archival literature.⁹ Treatises pointing out the shortcomings of traditional forms of quantitative inquiry, especially experimental research, appeared frequently in educational research journals.¹⁰ Much of the research that undergirds whole language comes from this more qualitative, more interpretive, more critical tradition. Thus, the credibility of this type of research increased in concert with the influence of whole language as a curricular movement.

Somewhere in the mid-1990s, the discourse of literacy research began to take a new turn. Stimulated by research supported by the National Institute for Child Health and Human Development, a "new" brand of experimental work began to appear, beginning in the mid-1980s and gathering momentum steadily since that time (Lyon, 1995; Lyon & Chhaba, 1996). This is experimentalism reborn from the 1950s and 1960s, with great emphasis placed on "reliable, replicable research," large samples, random assignment of treatments to teachers and/or schools, and tried and true outcome measures.¹¹ It finds its aegis in the experimental rhetoric of science and medicine and in the laboratory research that has examined reading as a perceptual process.¹² Although it was not broadly accepted by the reading education community when it first appeared, this work found a very sympathetic ear in the public policy arena.¹³

The political positioning of this research is important, but so is its substance. Two themes from this work have been particularly important in shaping a new set of instructional practices—phonemic awareness and phonics instruction.

The absolutely critical role played by phonemic awareness (the ability to segment the speech stream of a spoken word, e.g., /cat/ into component phonemes /cuh + ah + tuh/ and/or to blend separately heard sounds, e.g., /cuh + ah + tuh/ into a normally spoken word /cat/) in the development of the ability to decode and to read for meaning has been well documented in the past

decade and a half (Adams, 1990; Juel, 1988; Snow, Burns, & Griffin, 1998). Irrespective of mode of instruction, the overwhelming evidence suggests that phonemic awareness is a necessary but not a sufficient condition for the development of decoding and reading. First, children who possess high degrees of phonemic awareness in kindergarten or early in first grade are very likely to be good readers throughout their elementary school careers (Juel, 1988). Second, almost no children who are successful readers at the end of Grade 1 exhibit a low level of mastery of phonemic awareness. On the other hand, a substantial proportion of unsuccessful end-of-Grade-1 readers possess better than average phonemic awareness; this evidence is the critical piece in establishing that phonemic awareness is a necessary but not a sufficient condition for reading success. Although we can be confident of its critical role in learning to read, we are less sure about the optimal way to enhance its development. Many scholars have documented the efficacy of teaching it directly, but they also admit that it is highly likely to develop as a consequence of learning phonics, learning to read, or especially learning to write, especially when teachers encourage students to use invented spellings (see Adams, 1990; Juel, 1991). Research in whole-language classrooms (Clarke, 1988; Winsor & Pearson, 1992) suggests that writing is the medium through which both phonemic awareness and phonics knowledge develop—the former because students have to segment the speech stream of spoken words to focus on a phoneme and the latter because there is substantial transfer value from the focus on sound-symbol information in spelling to symbol-sound knowledge in reading.

The second consistent thread in the new experimentalism of the 1990s was the emphasis on the code in the early stages of learning to read. Reminiscent of Chall's (1967) earlier conclusions, scholars in this tradition advocated phonics first, fast, and simple.¹⁴ Less well documented, and surely less well agreed on, is the optimal course of instruction to facilitate phonics development. Even Gough (Gough & Hillinger,

1980), a classic bottom-up theorist, while arguing that what distinguishes the good reader from the poor reader is swift and accurate word identification, suggested that an early insistence on reading for meaning may be the best way to develop such decoding proficiency. Both Juel (1991) and Gough are convinced that students can learn how to read when they have *cryptoanalytic intent* (a disposition to decipher the specific letter-to-sound codes), phonemic awareness, an appreciation of the alphabetic principle (i.e., regardless of the numerous exceptions, letters do stand for sounds), and “data” (some texts to read and someone to assist when the going gets tough).

After reviewing available instructional evidence, two of the most respected scholars in this tradition, Marilyn Adams and Connie Juel, independently concluded that children can and should learn the “cipher” through a combination of explicit instruction in phonemic awareness and letter-sound correspondences, a steady insistence on invented spellings as the route to conventional spellings in writing activities, and lots of opportunity to read connected text (especially when the texts contain enough decodable words to allow students to apply the phonics information they are learning through explicit instruction). Both of these reviewers, known for their sympathies toward instruction in the code, are quick to add that rich experiences with language, environmental print, patterned stories, and “big books” should also be a staple of effective early reading instruction (Adams, 1990; Juel, 1991).¹⁵

This new research paradigm became officially codified by the appearance, in rapid succession, of two research syntheses—the publication of the report of the National Academy of Science's Committee on Preventing Reading Difficulties (Snow et al., 1998) and the report of the National Reading Panel (NRP) (2000). These are very different documents, and they have exerted very different influences on the reading field, particularly on reading policy. The *Preventing Reading Difficulties* report was conducted in the tradition of “best evidence” syntheses: well-established scholars meet, decide on

the issues, the domain of relevant research, and some subdivision of labor, do the work, write up the results, and turn the manuscript over to a set of editors to bring some synthetic clarity to the entire effort. As such, it considered a range of studies conducted within very different research traditions using very different research methods. The result was an apology for a balanced view of reading instruction, but with a special nod to phonemic awareness and phonics first and fast. A solid piece of scholarship many of us thought, but not much news (Pearson, 1999).

Authorized by congressional mandate, the NRP report used the most “scientific” review approaches (i.e., meta-analysis, at least wherever they could) available to them to distill from existing research what we knew about the efficacy of teaching phonemic awareness, phonics, fluency (instantiated as either guided reading instruction or independent reading), comprehension, and vocabulary; in addition, they investigated the status of the research base on teacher education and professional development and attempted to review research on technology and literacy. It is interesting to note that according to Catherine Snow (2001), one of the lead authors of the *Preventing Reading Difficulties* report, officials such as G. Reid Lyon and Duane Alexander from the National Institute for Child Health and Human Development, one of the sponsoring agencies of the NRP, were concerned about the *Preventing Reading Difficulties* report because it was vague and did not discriminate between trustworthy and untrustworthy research. The NRP report is noteworthy on a number of grounds. First, the actual conclusions in the main report are consistent with earlier attempts to summarize the knowledge base on these key issues, such as *Becoming a Nation of Readers* (Anderson, Hiebert, Scott, & Wilkinson, 1984) and *Preventing Reading Difficulties* (Snow et al., 1998), and point to a balanced approach to teaching reading. Second, although the vote of confidence in teaching phonics and phonemic awareness was strong and direct, it was moderated by important caveats that limit the applicability of these important instructional tools. For

example, phonics was found to be a useful instructional approach, but only in a particular time frame (Grades K–1); it was not effective for older students. Moreover, although the analysis privileged systematic phonics, nothing in the analysis implicated a particular approach (e.g., synthetic or letter-by-letter phonics vs. analytic phonics), nor was there any explicit support for decodable text. Also, the authors of the NRP report were careful, in their conclusions, to suggest that phonics by itself was not the total reading program: “Finally, it is important to emphasize that systematic phonics instruction should be integrated with other reading instruction to create a balanced reading program. Phonics instruction is never a total reading program” (NRP, 2000, p. 2-135).

Third, the authors of the NRP report were very clear about which topics and studies would be included. It would review only those topics for which there existed a sufficiently large pool of “potentially viable” experimental studies. Hence issues of grouping, the relationship of reading to writing, the role of texts in reading acquisition—just to name a few of the more obvious issues that schools and teachers must address in crafting local reading programs—are not addressed at all. Regarding specific studies, they would include only those that met minimal criteria: employ an experimental or quasi-experimental design with an identifiable comparison group, measure reading as an outcome, describe participants, interventions, study methods, and outcome measures in sufficient detail to “contribute to the validity of any conclusions drawn.” Natural experiments of the sort found in large-scale evaluation efforts or epidemiological investigations of relationships between methods and outcomes were excluded.

Vis-à-vis whole language, the point is straightforward: The changes in the dominant paradigm meant that the research base on which whole language was grounded (all of those close ethnographies of individual classrooms and teacher action stories) was no longer privileged in official conversations about “research”-based practice. Numbers, not compelling stories, were

the order of a new day; and it was not clear whether there was a place for constructivist pedagogy in general or whole language in particular, in these new conversations.

Politicization of the Reading Research and Policy Agenda

From its beginnings, one of the great hopes of educational research (and those who conduct it) has been that policy makers will take research seriously when they establish policy at a local, state, or national level. After all, the improvement of educational practice is the ultimate goal of educational research, and policy is our society's most transparent tool for educational improvement. Historically, however, research has been regarded as one among many information sources consulted in policy formation—including expert testimony from practitioners, information about school organization and finance, and evaluations of compelling cases. In the past half decade, research, at least selective bits of research, has never been taken more seriously. Several laws in California make direct references to research. For example, in 1998, California Assembly Bill 1086 prohibited the use of Goals 2000: Educate America Act of 1994 money for professional developers who advocated the use of context clues over phonics or supported the use of “inventive [sic] spellings” in children's writing. The federally sponsored Reading Excellence Act of 1998, which allocated U.S.\$240,000,000 for staff development in reading, required that both state and local applications for funding base their programs on research that meets scientifically rigorous standards. The *scientifically rigorous* phrase was a late entry; in all but the penultimate version of the bill, the phrase was *reliable, replicable research*, which had been interpreted as a code word for experimental research. As of early 1999, “phonics bills” (bills mandating either the use of phonics materials or some sort of teacher training to acquaint teachers with knowledge of the English sound-symbol system and its use in teaching) had been passed or were pending in 36 states.¹⁶

The NCLB made this goal of “evidence-based practice” even more explicit, with the phrase *scientifically based reading research* appearing more than 110 times in the Reading First portion of this act reauthorizing Title I.

Policy makers like to shroud mandates and initiatives in the rhetoric of science, and sometimes that practice results in strained, if not indefensible, extrapolations from research. This has happened consistently in the reading policy arena in the past decade. Three examples make the point vividly. First, California Assembly Bill 1086, with its prohibition on context clues and invented spelling, represents an ironic application of research to policy. The irony stems from the fact that many of the advocates of a return to code emphasis, such as Marilyn Adams, read the research as supporting the use of invented spellings in the development of phonemic awareness and phonics (Adams, 1990). Second, the mandate in several states calling for the use of decodable text (usually defined as text consisting of words that could be sounded out using a combination of the phonics rules taught up to that point in the program plus some instant recognition of a few highly frequent “sight” words) is based on the thinnest of research bases. The idea is that children will learn to use their phonics better, faster, and more efficiently if the texts they read permit facile application of the principles they are learning. Although it all sounds very logical, there is precious little research evidence to support the systematic and exclusive use of decodable text.¹⁷ This lack of evidence, however, does not seem to have deterred advocates who, on the phonics issues, championed scientific evidence as the gold standard for policy implementation.

The third example comes from the state of California's application for Reading First funds. The Reading First provision of NCLB requires that all elements of a program's application—instructional materials, assessments, and professional development—be supported by scientifically based reading research. *Scientifically based professional development* was defined in the California application as the professional development

required to help teachers implement the two state-adopted commercial reading programs; the proposal was accepted by federal officials without objection to this definition. This development was convenient in a financially troubled state that could ill afford to pay for the professional development for its new adoptions entirely on its own hook. The irony here, of course, is that the two commercially adopted programs in California, although they might be able to trace 15% to 20% of their practices to scientific research, are no more research based, let alone scientifically based (i.e., they have not regularly used randomized trials to test their efficacy) than the average run-of-the-mill commercial program. They are now, of course, officially blessed as scientific.

When research moves into the policy arena, one of two outcomes are most likely. If the research is widely accepted by members of the profession from which it comes, widespread acceptance and implementation usually follows. This often occurs in medical, pharmaceutical, or agricultural research. If widespread consensus on what the research says about practice is not reached, then research-based policy initiatives are likely to sharpen and deepen the schisms that already exist and the whole enterprise is likely to be regarded as a “war” among balkanized factions within the field. The latter scenario appears to characterize the reading field. The entry of science into the reading research community, and its accompanying blessing of particular approaches to teaching reading, has met with considerable resistance, some overt and some quiet, within the reading research community. The most vocal and prominent voices in the resistance have been Elaine Garan, Denny Taylor, and Richard Allington. Soon after the publication of the report of the Committee on Preventing Reading Difficulties of the National Academy of Science in 1998, D. Taylor (1998) published her treatise unveiling the “spin doctors of science.” Essentially, D. Taylor attempted to show how the conservatives involved in promoting the “new-phonics” agenda had used public relations techniques rather than science to

accomplish two goals: (a) to convince policy makers and the general public that the answer to teaching reading was more phonics earlier, and (b) to discredit public education more generally. Garan’s (2001, 2002) critique focused on the report of the NRP, and essentially, she offers two types of critique: internal and external. The internal critique holds the methodology of meta-analysis to its own standards, and she tried to show that the NRP effort was a fundamentally flawed approach to meta-analysis. For example, a principle of meta-analysis (Salkind, 2000) is that although the outcome measures need not be identical from one study to another, they should represent the same underlying construct; the NRP phonics analysis, Garan argued, fails this standard. She pointed out the many internal contradictions in method: for one group, eight studies are too few to move ahead with the meta-analysis whereas for another, nine is enough. However, perhaps most important, Garan pointed out that the statements included in the executive summary of the report are often inconsistent with comparable statements in the more elaborated reports of the various subgroups (on phonics, comprehension, and the like). I could not agree more with this last critique; as I will point out later, these discrepancies with the elaborated report only worsen when we examine the more “popular” version of the report written for general consumption and the headlines distilled by reporters for headlines and newspaper articles. Allington (2002) took a third approach. He enlisted the help of several colleagues in his edited volume to make the case that for the past 30 years, a conservative lobby has been trying to manipulate several policy levers (standards, assessment, professional development, and evidence-based practice) to shape a national reading policy that privileges basic skills for students and limits teacher education to training rather than educative practices. The case he made could be characterized as a sort of “skill the kids and de-skill the teachers” approach (my words, not his).

Interestingly, the debate, accompanied by its warlike metaphors, appears to have more life in the public and professional press than it does in

our schools. Reporters and scholars revel in keeping the debate alive and well, portraying clearly divided sides and detailing a host of differences of a philosophical, political, and pedagogical nature (see Manzo, 1997, 1998a, 1998b). Teachers, by contrast, often talk about, and more important enact, more balanced approaches. For example, several scholars, in documenting the practices of highly effective, highly regarded teachers, found that these exemplary teachers employed a wide array of practices, some of which appear decidedly whole language in character (e.g., process writing, literature groups, and contextualized skills practice) and some of which appear remarkably skills oriented (explicit phonics lessons, sight word practice, and comprehension strategy instruction). Exemplary teachers (e.g., Pressley et al., 2001; B. M. Taylor, Pearson, Clark, & Walpole, 2000; Wharton-MacDonald, Pressley, & Hampton, 1998) appear to find an easier path to balance than either scholars or policy pundits.

Producing Measurable Results

Evaluation has always posed a conundrum for whole-language supporters. First, some advocates oppose the use of any sort of externally mandated or administered assessments as a matter of principle, holding that assessment is ultimately the responsibility of a teacher in collaboration with a student and his or her parents. Second, even those supporters who are open to external forms of accountability, or at least reporting outside the boundaries of the classroom or school, often claim that standardized tests, state assessments, and other external measures of student accomplishment do not provide sensitive indicators of the goals of curricula based on whole-language principles. Most appealing would be assessments that are classroom based and individualized in nature, with the option of aggregating these sorts of data at the classroom and school levels when accountability comes knocking. During the 1990s, many felt that the increased emphasis on performance assessment and portfolios would fill this need.¹⁸

In an age of high expectations, explicit standards, and school- and classroom-level accountability, none of these options is a good fit with the views and desires of policy makers and the public. Both of these constituents seem quite uneasy about the quality of our schools and our educational system, so uneasy that leaving assessment in the hands of our teachers seems an unlikely outcome. It is not at all clear to me that the proponents of at least strong versions of whole language can, or will be willing to, hold themselves accountable to the sorts of measures that the public and policy makers find credible.

Loss of the Moral High Ground

One other factor, although difficult to document, seems to be operating in the rhetoric of the field in the first years of the 21st century. Whole language, and constructivist approaches generally, has always privileged the role of the teacher as the primary curriculum decision maker. Teachers, the argument goes, are in the best position to serve this important role because of their vast knowledge of language and literacy development, their skills as diagnosticians (they are expert “kid watchers”), and the materials and teaching strategies they have at their disposal. And in the arguments against more structured approaches, this is exactly the approach whole-language advocates have taken: “Don’t make these decisions at the state, district, or even the school level. Arm teachers with the professional prerogative (and corollary levels of professional knowledge) they need in order to craft unique decisions for individual children.” Although this may seem a reasonable, even admirable position, it has recently been turned into an apology for a self-serving teacher ideology.¹⁹ The counter argument suggests that the broad base of privilege accorded to teachers may come at the expense of students and their parents. Thus, those who advocate a strong phonics-first position often take the moral high ground: “We are doing this for America’s children (and for YOUR child!)—so that they have the right to read for themselves.” Even if one opposes this rhetorical move, it is

hard not to appreciate the clever repositioning on the part of those who want to return to more phonics and skills.

The Net Effect

Taken together, these factors created a policy environment in which whole language, or any other constructivist movement for that matter, was unlikely to flourish as the mainstream approach to teaching reading and writing. In the final analysis, however, I believe that the reluctance to own up to the “measurable results” standards was the Achilles’ heel of whole language. If whole-language advocates had been willing to play by the rules of external accountability, to assert that students who experience good instruction based on solid principles of progressive pedagogy will perform well on standardized tests and other standards of performance, they would have stood a better chance of gaining a sympathetic ear with the public and with policy makers. And as long as the criteria for what counts as evidence for growth and accomplishment are vague or left to individual teachers, the public could question the movement and wonder whose interests were being served by an unwillingness to commit to common standards.

LOOKING AHEAD

So where has this journey left us? And where will it take us next? I want to divide my analysis of the future of reading policy into two strands, research and curriculum, because these two faces of reading policy, although often joined at the hip, occasionally privilege different themes and issues. I will close by bringing them back together.

Research Policy

Complementarity as a Scientific Value

In the current research context, literacy scholars find themselves between a rock and a hard

place. The official views of research promulgated by the federal government in its research programs administered within the Department of Education are weighted toward quantitative and experimental work. At the same time, the work of many, perhaps even most, literacy researchers and doctoral students in research training programs is decidedly qualitative, narrative, and/or ethnographic in character. An impending crisis? A confrontation of the immovable object and the irresistible force? Or just the exclusion of a wide array of literacy scholars from federally funded research efforts? I would bet on the exclusion, but I hope and argue for a rapprochement among methods and even epistemologies.

Regarding science, my fundamental claim is that reading research can never be truly rigorous, indeed truly scientific, until and unless it privileges all of the empirical and theoretical methodologies that characterize the scientific disciplines. Included among those methodologies would surely be experimentation and of course randomized field trials of the sort that are being proposed for several federally sponsored programs, but the range of scientific methods would extend to:

- careful descriptions of phenomena in their natural settings (just like Darwin did and just like today’s environmental scientists);
- examinations of natural correlations among variables in an environment, just to see what goes with what;
- natural experiments in which we take advantage of the differences that serendipity and the normal course of events have created between two or more settings that are otherwise remarkably similar—the most common form of this effort in education being outlier studies and the even more common approach in public health’s epidemiological studies;
- data gathered in the name of theory building and evaluation—just to see if we can explain the nature of things;
- design experiments in which we adopt a planful, incremental approach to knowledge refinement, with each successive step building carefully on what was learned in the last; and

- the use of qualitative tools such as ethnography and discourse analysis in concert with randomized experiments to describe what is really going on inside those randomly assigned treatments, so that we can explain why a treatment worked or did not work, or whether the range of variation in treatments is so great across sites that it is doubtful that it can really be called the same intervention across sites, or what the consequences, especially the unintended consequences, of an intervention might be.

As good as randomized experiments are for determining the overall efficacy of interventions, they are very short on details about the interventions, such as why, how, for whom, and under what conditions interventions work. For that we need complementary methods, and this is where qualitative methods come into play. Donald Campbell (1984), one of the foremost design methodologists of the 20th century and the coauthor of the infamous book on quasi experiments (Campbell & Stanley, 1963), the classic treatment of threats to internal and external validity, recognized this need for complementarity:

To rule out plausible rival hypotheses we need situation-specific wisdom. The lack of this knowledge (whether it be called ethnography, program history, or gossip) makes us incompetent estimators of program impacts, turning out conclusions that are not only wrong, but are often wrong in socially destructive ways. . . .

There is the mistaken belief that quantitative measures replace qualitative knowledge. Instead, qualitative knowing is absolutely essential as a prerequisite for quantification in any science. Without competence at the qualitative level, one's computer printout is misleading or meaningless. (pp. 141–142)

We hear a lot of talk about randomized field trials in medical and pharmaceutical research, and we are advised to follow their lead. I agree. But if we follow medicine and pharmacology, then we should follow them all the way down the road of science. Let us remember that before researchers in those fields get to the last 10% of the journey, which is when they invoke randomized field trials in anticipation of advocacy and

policy recommendations, they have already used a much wider range of methodologies, including much observation, description, examinations of relationships, and just plain messing around (that is a technical term used by scientists to describe what they spend most of their time doing) to travel the first 90% of that journey. So let us talk about complementarities and convergence among methods rather than competition and displacement of one worldview with another. This is the message of the recent report on educational research by a committee empanelled by the National Academy of Science (Shavelson & Towne, 2002), a message I heartily endorse.

If we rush too soon to the last 10% of the journey and enamor ourselves of randomized field trials for their own sake, we are likely to end up conducting expensive experiments on interventions that were not worth evaluating in the first place. A drug company would never think of conducting a randomized field trial on a new drug that had not gone through a thorough basic research phase in which biochemical theories, tryouts on nonhuman organisms, correlational research on chemical components of the drug in the natural environment, and probably some serendipitous case studies of individual subjects who volunteered to use the drug out of desperation all played a key role. We should ask no less of educational interventions and programs. An intervention that is based on bad theory or no theory is not likely to yield a significant contribution to practice in the long run. To know that something worked without a clue about how and why it worked does not advance either our scientific or professional understanding of an educational issue. We cannot afford blind experimentation and horse races with interventions of unknown theoretical characteristics. As our candidates for randomized field trials, we want treatments and interventions that have gone through these various stages of scientific development.

I fear that as a profession we have fallen into a methodological trap. We have become so attached to our methodologies and to their epistemological (some would say ideological) underbellies that we, as individuals, are likely to begin our work by

looking for a question that fits our methodological preferences rather than the other way around. This does not serve our profession well, for it allows us to address questions that may or may not be of great relevance to policy and practice. We must return to the ethic of insisting that just as form follows function in language, so methods must follow questions in research. And if we do not, as individuals, possess the range of methodological expertise to address different sorts of questions, then we ought to align ourselves with scholarly communities in which such expertise is distributed among its members.

As a curious and ironic footnote, I would point out that complementarity across methods is consistent with the definition of *scientifically based reading research* in the Reading First portion of NCLB; the definition includes these standards:

- employs systematic, empirical methods that draw on observation and experiment;
- involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions;
- relies on measures that provide valid data across observers and occasions; and
- is published in peer-reviewed journals (or reviewed by a duly constituted panel).

The Complexity of Research in Education

Complexity in the policy arena is always a double-edged sword. To assert that educational research is complex is to imply that there is something categorically different about educational research in comparison to research in agriculture, physics, chemistry, medicine, or even psychology. Usually the complexity is attributed to the human factor and the variation introduced by human activity:

- that individuals differ from one another;
- that they live and work in groups;

- that the members of the group influence what others do, how they act, and what they believe; and
- that when humans are involved, things change in unpredictable ways.

David Berliner (2002), in a persuasive article in a recent issue of *Educational Researcher*, puts forward just such a view. And there is much truth in the argument. I know this all too well from my own experience in trying to do large-scale research on best practices (B. M. Taylor, Pearson, Peterson, & Rodriguez, 2003). We tried to do an outlier analysis of schools that beat the odds predicted by their demographics. However, we found that high poverty/high performance status is not a static characteristic. Some schools that entered the study with a record of high achievement foundered; some with reputations as failing the mark changed their ways. Had we not collected a wide array of student outcomes (which allowed us to build post hoc indicators of who was and was not beating the odds) and an even wider array of indicators of school reform efforts and teacher practices, we would not have been able to unearth program and instructional characteristics that explain variation in achievement growth. Moreover, we found that a combination of quantitative and qualitative approaches were absolutely essential in teasing out important relationships between programs and outcomes.

We continue to find, in our more recent work (B. M. Taylor et al., 2003) with low-income, low-performing, aspiring schools, that things are not always what they appear to be—that there is incredible variability among our intervention schools in the degree to which the intervention is actually implemented, both across schools and across classrooms within schools. We also find that the degree of fidelity to the intervention, not to a set of specific instructional practices but to a set of broad principles outlining the process to be followed and the issues to be addressed, is a good predictor of achievement growth, again both across schools and across classrooms within schools. My point is simple—no matter how well planned an intervention might be, things happen

and variation will occur. In many studies, the variation within treatments is often equal to the variation between treatments. Thus, it is critical in research involving programmatic and instructional reform, to document carefully the nature of the actual practices across schools and classrooms. And there is no better tool to do this than ethnographic descriptions of classroom instruction and professional development meetings. In short, we need all the tools we can muster to address the inherent complexity of research involving human beings who live and work in groups.

The final question about randomized field trials is whether we will be willing to pay the price tag. It is one thing to randomly assign college freshmen who happen to have the misfortune to be enrolled in Psychology 1-A to different treatments. It is quite another to randomly assign teachers, classrooms, and even schools to a particular treatment. In the psychology class, I test 30 subjects and I get 30 data points for my analysis. In the classroom, I test 30 students and get one data point—the classroom mean. That is one cost factor. But there are others: For example, if we want to know if the treatment generalizes across types of students and types of schools, then we will either have to draw very large samples or very carefully shaped selective samples.

*The Treacherous Road
From Research to Policy*

The road from research to policy is fraught with many dangers—potholes, blind corners, road hogs, and detours that can frustrate even the most thoughtful traveler. Both researchers and policy makers must be aware of these threats as they do their best to draw valid inferences from research for practice and policy. Let me unpack some of the dangers and some guidelines for minimizing risk to students, parents, and teachers.

When research travels to the land of policy, often only the headlines make the journey, leaving the details and the nuance behind. The consequences of this fact of policy life are depicted with real examples of the discrepancy in Table 1.1.

I am not sure the journalists are to blame; the reporting of educational research probably does not differ much from the reporting of medical, pharmaceutical, agricultural, or public health research. Lest we think that education is different, compare these two headlines, one composed by a staff writer of the *New York Times* and the other by a staff writer of the *Washington Post* in reporting the findings from an article released by the *New England Journal of Medicine* about the relative effectiveness of surgery versus benign neglect in treating prostate cancer in men:²⁰

From the *New York Times* on September 12, 2002: “Prostate Cancer Surgery Found to Cut Death Risk” (Kolata, 2002)

From the *Washington Post* on September 12, 2002: “Prostate Cancer Therapies About Equal” (D. Brown, 2002)

One glass is half full; the other half empty. A person contemplating surgery would much rather be reading the *Times*! The point is that as a society we must find a way to cope with the persistent problem of interpretation that tends toward oversimplification, whether it occurs in the press, the Congress, or our statehouses. Perhaps we should require that policy makers (or members of their staffs) be required to read beyond the headlines of educational reports before setting policy in concrete. Nuance may not make things simple, but nuance is a fact of life in most policy contexts, including public health and medicine.

Research is often used in a selective, uneven, and opportunistic manner by policy makers. An unfortunate corollary of this surface-level approach to summarizing research for policy purposes is the uneven use of the research card in setting policy. The danger is that those who set policy will choose to play that card when the evidence swings in their favor; and when it does not, they will appeal to common sense, the conventional wisdom of practice, or authoritative opinion. So, for example, when the NRP report blessed the systematic teaching of phonics,

Table 1.1 Headlines Versus Details in the Reporting of Reading Research

<i>Headline</i>	<i>Source</i>	<i>Digging Deeper Into the Actual Report</i>
Systematic, Explicit, Synthetic Phonics Improves Reading Achievement	Foorman, Francis, Fletcher, Schatschneider, and Mehta (1998)	When a program includes systematic, synthetic phonics among many other elements (lots of writing, lots of reading of a whole range of texts, and lots of supplementary activities), a small but robust effect for a subset of the population is found on a measure that requires kids to read lists of pseudowords.
Phonemic Awareness Improves Later Reading Achievement	National Reading Panel (2000)	Phonemic awareness helps . . . <ul style="list-style-type: none"> • If taught early (K–1); • Mostly on measures of word identification; • If taught with letter-sound instruction; • If limited in scope (from 18 to 20 hours).
Phonics Wins	National Reading Panel (2000)	Phonics helps . . . <ul style="list-style-type: none"> • If it is taught early (not great beyond Grade 1); • More on word recognition than comprehension; • If it is systematic and explicit (no evidence for one approach over another); • If it is embedded in a rich curriculum; • If caveats are recognized, for example, that there is no evidence for decodable text.
Independent Reading Does Not Help—If You Want to Do It, Assign It as Homework	National Reading Panel (2000)	The National Reading Panel did not study independent reading but rather the impact on fluency of instructional interventions designed to increase the amount of independent reading done in classrooms. From the paltry array of studies they were able to assemble, they concluded that the research on the efficacy of such interventions was inconclusive.

basic-skills advocates were quick to point to the scientific evidence underlying their policy initiatives. But the use of decodable text in conjunction with those programs, which could not be justified by the available evidence, was rationalized as a commonsense adjunct to a systematic approach to teaching phonics. In a similar vein, the Los Angeles Unified School District has

mandated that all high schools use a remedial program with the ironic title of *Language* for all of its low-performing secondary students. *Language* is a decodable text (“Dan can fan Nan”) throwback to the linguistic readers of the mid-1960s. The noteworthy aspect of its adoption is it was adopted not by appealing to the evidence (the NRP could not document the use of

phonics for readers in that age range) but by appealing to common sense (these kids clearly missed out on phonics the first time around so let's go back to square one and do it right). Allington and Woodside-Jiron (1998) have documented similar enactments of selective attention to research in noting the widespread adoption of decodable texts in state textbook standards.

Some science is more important than other science. Another corollary of this uneven use of research is a kind of first among equals conspiracy of good intentions. And it applies to the use of the NRP report in setting policy. The chapters on comprehension and vocabulary in the NRP are laudatory in their praise for the work in these areas (although they eschewed meta-analysis in favor of a best evidence synthesis on the grounds of too few studies) and enthusiastic in recommendations for renewed attention to strategy instruction and ambitious vocabulary teaching. Moreover, the even more recent Rand report (Snow, 2002) advocates a renaissance in research in comprehension instruction and assessment. But I have not witnessed a groundswell in advocacy for comprehension and vocabulary instruction as the fundamental solution to America's literacy problems. There seems to be a kind of "first things first" ethic, suggesting that "of course we'll get to comprehension and vocabulary . . . , but first let's make sure we have the basics in place." The same could be said for the section on teacher education and professional development in the comprehension chapter of NRP; we get glowing recommendations for the efficacy of professional-education models to increase teacher capacity to teach comprehension but little action on the policy and "scientifically based" professional development fronts. One of the other ironies of professional development in NCLB is that what it means to conduct evidence-based professional development is that the content of the professional development sessions must be based on scientifically based reading research about how young children learn to read, but need not attend at all to the substantial body of research documenting the optimal ways

to promote teacher learning. We know from a host of studies (see Richardson & Placier, 2002; Wilson & Berne, 1999) that professional learning is at its best when teachers have a voice in its design, when it is long term and school based, when it is focused on analyses of teaching and student learning, and most important, when the focus is on establishing learning-sustained communities. Yet NCLB, as committed as it is to science, is moot on the point of how professional development (itself to be based on scientifically based reading research) ought to be organized or delivered.

When we do not have definitive research to answer a question about policy or practice, we can easily slip over the line and privilege ideology and belief over evidence. The gold standard in research leading to policy implementation is surely the randomized field trial (see Mosteller & Baruch, 2002, for a series of articles extolling the virtues and assessing the limitations of the randomized field trial). Other things being equal, it is better to have experimental evidence to support a claim or validate a practice. But what are educators, especially school- and district-based educators, to do when they must establish curricular practices or forge new programs without the benefit of randomized experiments? Do we go straight from randomized trials to personal beliefs? Or do we establish a principle that requires us to use the best evidence available to us in any situation? If there are no randomized field trials, can we rely on the evidence from quasi experiments? Natural experiments? Correlational research? Best-practice research? They are all quantitative, but they cannot easily rule out rival hypotheses. How about case studies? Ethnographies? They are both empirical- and data-driven but they have a different set of evidentiary rules and a different notion of generalizability (Firestone, 1987). And when do we resort to professional consensus, the wisdom of experience, and personal belief? As a profession, we have not established any sort of hierarchy of evidentiary sources (Peirce, 1885, as cited in Hartshorne, Weiss, & Burks, 1931–1958).

Perhaps it is time we did. It is my personal conviction that it is our moral and ethical obligation to use the best evidence we can muster for making policy decisions of consequence. Further, when we have no evidence, we must fess up to that fact and make it clear that we are basing policy on values, beliefs, and hunches. If we followed some sort of evidentiary guidelines, we would not have so many intensive phonics programs for older students, so much decodable text in our commercial programs, or so little time for independent reading.

The independent reading issue is particularly troubling because it took a double hit. First, the NRP chose to examine it even though there were precious few studies that could pass through the eye of the needle imposed by their standards for inclusion. When they did, they stated their conclusion in a way that allowed readers to move from what they did say, “There is no evidence to support the efficacy of school-based programs that promote independent reading,” to what some wanted to hear, “Independent reading is a waste of time.”

Second, the NRP did not look at anything but experiments, thus eliminating some powerful evidence documenting the importance of everyday reading. Had the NRP examined a wider array of research, including several experiments conducted in other countries for second-language learners (e.g., Elley, 1998), a few smaller scale experiments (e.g., B. Taylor, Frye, & Maruyama, 1990), some impressive naturalistic studies conducted in an epidemiological tradition (e.g., Anderson, Wilson, & Fielding, 1988), and a wide array of best-practice research (e.g., B. M. Taylor, Pearson, et al., 2000; Wharton-MacDonald et al., 1998), they would have reached a very different conclusion about the efficacy of just plain reading. And in a situation in which the experimental research is moot, that would have been the high road to take on such a key professional issue.

Curricular Policy

Many recent developments suggest that we are retreating to a more familiar, more comfortable

paradigm of basic skills in which phonics, skills, and controlled text dominate our practices. Other developments suggest that we are on the verge of a new paradigm, a hybrid that weds some of the principles of whole language (integrated instruction and authentic texts and tasks) with some of the traditions of earlier eras (explicit attention to skills and strategies, some vocabulary control of early readers, and lots of early emphasis on the code) in an “ecologically balanced” approach to reading instruction.²¹ The most cynical among us might even argue that we are just riding the natural swing of a pendulum that will, if we have the patience, take us back to whole language, or whatever its pedagogically constructivist, child-centered descendant turns out to be, in a decade or so. Before making a prediction about the direction the field will take, let me play out the first two scenarios: phonics first and balanced reading instruction.

Two Different Worlds

If those who have advocated most strongly for a return to phonics and a heavy skills orientation have their way—if they are able to influence federal, state, and local policy as well as the educational publishing industry—we will experience even bigger shifts in the very earliest stages of learning to read—preschool, kindergarten, and Grade 1. They suggest explicit instruction on phonemic awareness and phonics, with a strong preference for decodable texts in the early grades. When it comes to writing, literature, response, and comprehension, the phonics-first advocates seem quite content to cede curricular authority to the practices that emerged during the 1980s and early 1990s, those associated with whole language, literature-based reading, and process writing (see Adams & Bruck, 1995; Fletcher & Lyon, 1998). Thus, looking broadly at the entire elementary reading curriculum (the range of materials and the range of pedagogical practices), things might on the surface look similar to the early 1990s, with some retreat to the 1980s, especially in terms of skill and strategy instruction.

But beneath that curricular surface, major changes would have occurred. For example, the role of the teacher and the learner would have reverted to what they were before the ascendancy of constructivist teaching reforms, such as whole language. The role of the teacher would be to transmit the received knowledge of the field, as reflected in research-based curricular mandates, to students. Students would eventually be regarded as active meaning makers, but only after they had received the tools of decoding from their teachers. The greatest changes of all would have taken place in the underlying model of reading and reading acquisition. The simple view of reading (that reading comprehension is the simple product of decoding prowess and listening comprehension) would have returned in full force, and the job of young readers would be to acquire the decoding knowledge they lack when they begin to learn to read.

If those who are pushing for ecological balance carry the day, the field will experience less dramatic shifts. A balanced approach will privilege authentic texts and tasks, a heavy emphasis on writing, literature, response, and comprehension, but it will also call for an ambitious program of explicit instruction for phonics, word identification, comprehension, spelling, and writing. A balanced approach is likely to look like some instantiations of whole language from the early 1990s, but recalibrated to redress the unintended curricular consequences outlined earlier in this chapter. Major differences between a balanced approach and the new phonics are likely to manifest themselves most vividly in kindergarten and Grade 1, where a rich set of language and literacy experiences would provide the context from which teachers would carve out scaffolded instructional activities to spotlight necessary skills and strategies—phonemic awareness, letter-sound knowledge, concepts of print, and conceptual development. Thus instruction, although focused and explicit, would retain the highly contextualized patina of whole language.

Beneath the curricular surface, balanced approaches seem to share slightly more in common, at least on a philosophical plane, with

whole-language than with new-phonics approaches. The teacher is both facilitator and instructor. The teacher facilitates learning by establishing authentic activities, intervening where necessary to provide the scaffolding and explicit instruction required to help students take the next step toward independence. The student is, as in whole language, an active meaning maker from day one of preschool. Reading is a process of constructing meaning in response to texts encountered in a specific context, and the emergent literacy metaphor, not the readiness metaphor, characterizes the acquisition process.

An Ecologically Balanced Approach

Just in case my personal bias has not emerged, let me declare it unequivocally. I favor the conceptual map of the ecologically balanced approach, both for research and curricular policy.

I hope my reasons for supporting ecological balance—or as Howe and Eisenhart (1990) have characterized it, *compatabalism*—in research methods are transparent. The problems we face are too vexing to limit ourselves to a single methodology or epistemology. Multiplicity in the tradition of Spiro's (Spiro & Jehng, 1990) cognitive flexibility theory is what is needed now. We surely need to know what works, but we also need to know why it works, for whom, and under what conditions; interestingly, this sort of approach is appealing to many research policy leaders, including those who have led the charge toward more experimental approaches (see Lyon, 2003; Whitehurst, 2001). For example, in testifying to Congress, Lyon (1999) expressed just the sort of complementarity I have argued for:

In order to develop the most effective instructional approaches and interventions, we must clearly define what works, the conditions under which it works, and what may not be helpful. This requires a thoughtful integration of experimental, quasi-experimental, and qualitative/descriptive methodologies.

To paraphrase Spiro and his colleagues, neither simplemindedness nor muddleheadedness will serve our interests, or those of the nation, well.

There are several reasons for favoring an ecologically balanced, or comprehensive, stance toward curriculum. First, my reading of the reading research points to the balanced-curricular position, not to the new-phonics position or the whole-language position, and it does so on both a theoretical and a pedagogical plane. I do not see much support for the simple view of reading that underlies the new phonics; readers do construct meaning, they do not just find it lying there in the text. Regarding pedagogical research, my reading requires me to side with Chall's (1967) view that although some sort of early, focused, and systematic emphasis on the code is called for, no particular approach can be singled out. Even the recent report of the NRP (2000) took exactly that position. And although I readily accept the findings of the phonemic awareness research, I do not read them as supporting drill and practice approaches to this important linguistic understanding; to the contrary, highly embedded approaches, such as invented spelling, are equally as strongly implicated in the research (see Clarke, 1988; NRP, 2000; Winsor & Pearson, 1992).

Second, an ecologically balanced approach is more respectful of the entire range of research in

our field. It does not have to exclude major research paradigms or methodological approaches to sustain its integrity.

Third, an ecologically balanced approach also respects the wisdom of practice. It is no accident that studies of exemplary teachers, those who are respected by their peers and nurture high student achievement, consistently find that they exhibit a balanced repertoire of instructional strategies. Teachers who are faced with the variations in achievement, experience, and aptitude found in today's classrooms need, and deserve, a full toolbox of pedagogical practices.

Finally, an ecologically balanced approach respects our professional history. It retains the practices that have proved useful from each era but transforms and extends them, rendering them more effective, more useful, and more supportive of teachers and students. And it may represent our only alternative to the pendulum-swing view of our pedagogical history that seems to have plagued the field of reading for most of the 20th century. A transformative rather than a cyclical view of progress would be a nice start for a new century.

