Research and the Reading Wars

James S. Kim

Controversy over the role of phonics in reading instruction has persisted for over 100 years, making the reading wars seem like an inevitable fact of American history. In the mid-nineteenth century, Horace Mann, the secretary of the Massachusetts Board of Education, railed against the teaching of the alphabetic code—the idea that letters represented sounds—as an impediment to reading for meaning. Mann excoriated the letters of the alphabet as "bloodless, ghostly apparitions," and argued that children should first learn to read whole words. The 1886 publication of James Cattell's pioneering eye movement study showed that adults perceived words more rapidly than letters, providing an ostensibly scientific basis for Mann's assertions.² In the twentieth century, state education officials like Mann have continued to voice strong opinions about reading policy and practice, aiding the rapid implementation of whole language-inspired curriculum frameworks and texts during the late 1980s. And scientists like Cattell have shed light on the processes underlying skillful reading, contributing to a growing scientific consensus that culminated in the 2000 National Reading Panel report.³

This chapter traces the history of the reading wars in both the political arena and the scientific community. The narrative is organized into three sections. The first offers the history of reading research in the 1950s, when the "conventional wisdom" in reading was established by acclaimed leaders in the field like William Gray, who encouraged teachers to instruct children how to read whole words while avoiding isolated phonics drills. In the 1960s and 1970s, Jeanne Chall's research on first-grade reading instruction

indicated that phonics instruction was effective in helping children become skilled readers. Whole-language theorists, however, conducted research to challenge Chall's findings, arguing that context clues helped children read more effectively than teacher-directed phonics instruction.

The next section shows how both the federal government and state education agencies mediated the reading wars. In the 1980s and 1990s, the federal government turned to experts to undertake syntheses of research and to highlight areas of scientific consensus, which could form a solid foundation for improving instruction. At the same time, advocates of whole-language pedagogy persuaded numerous states to adopt new curriculum frameworks and instructional materials that pushed phonics instruction to the periphery of the classroom. Eventually, the decline in fourth-grade readings scores on the National Assessment of Educational Progress (NAEP) prompted many state legislatures to pass mandatory phonics bills.

The third section discusses how findings from the National Reading Panel (NRP) synthesized three decades of scientific research in reading and formed the basis for the Reading First legislation. Since 2000, the NRP's findings have continued to fuel the ongoing debate about evidence-based practice in reading among state and federal policymakers, professional organizations, and teachers. In sum, this chapter seeks to describe *how* research has helped resolve controversies in the reading wars; and to explain *why* good research alone cannot ensure sound instructional policy and practice.

1967-1979: THE GREAT DEBATE IN EARLY READING INSTRUCTION

Phonics versus Look-Say: And the Winner Is?

In the mid-twentieth century, the conventional wisdom about effective reading instruction in the early grades was heavily influenced by William S. Gray, a leading reading scholar. In his 1948 book, *On Their Own Reading*, Gray objected to the dominant method of teaching children the letters of the alphabet, the sounds represented by letters, and the blending of groups of letters to sound out words. Echoing Mann's earlier criticism of phonics instruction, Gray objected to "the old mechanical phonic drills... that inevitably result in dull, word-by-word reading." Instead, he endorsed a meaning-first and word-analysis-later approach that became a hallmark characteristic of the classroom texts used to instruct children in reading. Gray and his supporters theorized that reading skill would develop more rapidly if children learned to look at and quickly recognize whole words much like adults. This method of instruction was called "look-say" because it taught children

to recognize and say whole words by sight rather than using knowledge of letter-sound relationships to read words.

In Why Johnny Can't Read, Rudolph Flesch attacked the basic premise of a meaning-focused instructional approach proposed by Gray. Flesch exhorted teachers and parents to instruct children how letters represented sounds and how to blend those sounds to identify unknown words. According to Flesch, phonics was the best way to teach reading, and the only hope for curing our nation's reading woes. Furthermore, Flesch directed his message to parents, teachers, and the general public, seeking to win the debate about how best to teach children to read.⁵ In the context of the cold war, Flesch's back-to-basics, phonics-first, message was embraced by many politicians and citizens who feared that the American educational system was losing ground to the Russians.⁶

With the support of a Carnegie Corporation grant, Jeanne Chall, a professor at Harvard University's Graduate School of Education, undertook a research synthesis to assess the competing ideas about early reading instruction. The title of Chall's 1967 book, Learning to Read: The Great Debate, captured the essence of the reading wars that erupted in the middle of the twentieth century. Chall noted that the many issues and controversies about reading instruction in first grade boiled down to one question: "Do children learn better with a beginning method that stresses meaning or with one that stresses learning the code?"7 To address this question, Chall interviewed teachers, inspected basal texts, and reviewed earlier research on elementary school reading. However, the heart of Chall's analysis was an assessment of the efficacy of three instructional approaches, based on a methodological strategy for aggregating findings from experimental studies. The looksay method emphasizes the visual recognition of whole words, the reading of whole sentences, and the acquisition of meaning; it involves little or no phonics instruction. The systematic phonics program teaches phonics early and instructs children how to separate letter-sound relationships and to blend these sounds. The intrinsic phonics program stresses sight-word reading, teaches children how to learn the sounds of letters by analyzing sight words, and encourages children to use context clues and pictures to identify words. Chall's classification system placed programs on a continuum from instruction emphasizing the code (systematic phonics) to programs emphasizing meaning (look-say) and the programs in-between (intrinsic phonics). Finally, Chall tallied the number of experimental studies favoring each of the three approaches to beginning reading.

Consulting research from 1900 to 1965, Chall reviewed 30 experimental studies that compared at least two different approaches to beginning read-

ing instruction. As shown in Table 4.1, 27 studies showed superior outcomes in phonics programs (18 favored systematic phonics and 9 favored intrinsic phonics) and 3 favored look-say methods. Contrary to the conventional wisdom, Chall found that an early code emphasis produced better word recognition outcomes in the early grades and helped children read with better comprehension up to fourth grade, relative to the dominant look-say method of reading instruction, in which little phonics was taught and emphasis was placed on reading whole words and whole sentences. A code-emphasis also produced larger benefits for less-skilled readers and children from low-income families. However, Chall emphasized that the dichotomy between code- and meaning-emphasis was a simplistic dualism and a matter of emphasis—all programs had some dose of instruction about letter–sound relationships and whole-word reading. Because the findings were based on few quality experiments, Chall viewed her findings as "hypotheses to be tested further."

A Psycholinguistic Theory of Reading: Theoretical Roots of Whole-language Pedagogy

Chall's findings were immediately challenged by two scholars—Kenneth S. Goodman and Frank Smith. In a 1967 journal article, Goodman challenged the idea that reading involved the "exact, detailed, sequential perception and identification of letters, words, spelling patterns and large language units."9 To Goodman, reading was a "psycholinguistic guessing game," in which good readers used context and background information rather than precise identification of letter-sound relationships to predict, confirm, and guess at the identification of an unfamiliar word. 10 In a 1969 article in Reading Research Quarterly, Goodman went on to criticize "recent attempts by Chall and others to justify the separation of code-breaking from reading for meaning."11 According to Goodman, syntax (grammatical structure of language) and semantics (relevant background knowledge) were as important as grapho-phonetic cues (letter-sound correspondences) in aiding word recognition processes. In the conclusion, Goodman hoped that his research would "generate hypotheses about the reading process which can be empirically tested and lead to new insights into methods and materials for reading instruction."12

As he hoped, Goodman's theory fueled research on the processes underlying word recognition ability among linguists, cognitive psychologists, and educational researchers. In the 1970s, Goodman's ideas were echoed most prominently in the writings of Frank Smith, who began to publish research originating from the Harvard University Center for Cognitive Studies under the direction of the eminent psychologist George Miller. In *Understanding*

TABLE 4.1 A Comparison of Three Approaches to Reading Instruction, 1900-1965

Period	Prevailing method	Total studies	Found systematic phonics superior	Found intrinsic phonics superior	Found look-say superior
1900–1920	Systematic phonics	3	1	0	2
1920-1935	Look-say	6	4	2	0
1935–1955	Intrinsic phonics	8	4	3	1
1955-1965	The debate: intrinsic phonics still the prevailing method with a push toward earlier and heavier emphasis on phonics	13	9	4	o
TOTAL		30	18	9	3

Source: Jeanne Chall, Learning to Reading: The Great Debate (New York: McGraw-Hill, 1967), 132.

Reading, Smith hypothesized that reading was natural, like speaking, and that children knew a good deal about language. Consequently, skilled readers used their knowledge of a meaningful context, like a story or paragraph, to recognize individual words in a sentence. Furthermore, Smith argued that phonics rules were too complex, had too many exceptions, and were unnecessary for helping the beginning reader become a proficient reader. "Learning to read," Smith argued, "is akin to any other skill; there are perhaps some specialized exercises that one can undertake to iron out particular difficulties, but there is no substitute for the activity itself." In other words, one of the best ways to become a skilled reader was simply to read.

In trying to bridge the divide between theory and practice, Smith directed his message to practitioners. The final section of *Understanding Reading* was titled The Teacher's Role, and Smith acknowledged that his model of reading was unconventional. He explained,

[Readers] are not usually regarded as "predicting" their way through a passage of text, eliminating some alternatives in advance on the basis of their knowledge of the redundancy of language, and acquiring just enough visual information to eliminate the alternatives remaining . . . But nothing that I have said should start a classroom revolution. There is no suggestion that teachers of reading should throw away their instructional procedures, or their years or experience, and start all over again. ¹⁵

JAMES S. KIM

The direct appeal to teachers and the deference to practitioner judgment became a theme in Smith's writing throughout the 1970s, and both Goodman and Smith started to make explicit instructional recommendations based on their theories about reading.

Smith published two books in the 1970s that were squarely focused on persuading teachers to change instruction in their classrooms. In Psycholinguistics and Reading, Smith reiterated his basic hypothesis that the primary goal of reading was to acquire meaning from print and dismissed the importance of teaching the systematic relationship between letters and sounds. Pre-packaged phonics programs, Smith claimed, undermined teacher autonomy, Indeed, Smith argued that teachers "do not act as brainless purveyors of predigested instruction (that is why there is the frightening trend these days to produce 'teacher-proof' materials)."16 In Reading Without Nonsense, Smith argued that phonics instruction was unnecessary for efficient decoding of words and actually interfered with the process of learning to read. When confronted with a struggling reader, Smith advised teachers: "The first alternative and preference is-to skip over the puzzling word. The second alternative is to guess what the unknown word might be. And the final and least preferred alternative is to sound the word out. Phonics, in other words, comes last."17

By the end of the 1970s, Smith's writings reflected the actions of a policy entrepreneur seeking to persuade teachers to reject phonics instruction and to encourage children to use context clues to identify words. Moreover, his recommendations evoked earlier arguments put forth by William Gray, who suggested that teaching children to read whole words was superior to phonics instruction.

Testing the Psycholinguistic Theory of Reading in Laboratory Experiments

Starting in the early 1970s, cognitive psychologist Keith Stanovich notes that Smith's top-down model of reading piqued the curiosity of scholars, who began to undertake experiments to examine whether good readers, in fact, relied more on context to recognize words than poor readers. The more immediate goal of these studies was to understand the basic processes underlying reading rather than the application of these findings to schools and classrooms. Much to the surprise of Stanovich and his colleagues, experimental data indicated that it was poor readers, not good readers, who relied more heavily on context to facilitate word recognition. Working independently on related questions, other scholars replicated the finding that context effects were largest for poor readers. Studies based on eye-movement tech-

nology also indicated that good readers did not engage in whole-scale skipping of letters and words but processed all the visual information in text.²¹ Mounting evidence also indicated that children needed to develop phonemic awareness—the knowledge that words are composed of units of sounds (phonemes) represented by letters (graphemes)—and master the alphabetic principle in order to become independent readers.²² Contrary to Goodman and Smith's top-down model of reading, research findings also showed that good readers had fast and accurate word recognition ability that freed attention to focus on meaning, whereas poor readers had to rely more heavily on context to decode words.²³

Although basic research produced knowledge about the acquisition of reading skills among young children, these findings were disseminated primarily in peer-reviewed journals and academic conferences and did not have an immediate and direct impact on education policy and practice. In assessing the state of reading research in 1977, Richard Venezsky argued, that "[if] reading research is to influence instruction, then more experimental psychologists will have to be persuaded to interact professionally with educational planners and developers and to concern themselves with the practical side of reading."²⁴ Venezsky's recommendations for using research to shape practice took two forms in the 1980s. The federal government convened expert panels to synthesize basic research and its implications for teachers, while whole-language advocates worked with state education agencies to disseminate their ideas about reading instruction to teachers.

1983–1997: HOW CONSENSUS PANELS OF EXPERTS AND STATE EDUCATION AGENCIES MEDIATED THE READING WARS

Consensus Panels and the Call for Balanced Literacy Instruction

In 1983, the National Institute of Education (NIE) authorized the National Academy of Education's Commission on Education and Public Policy to gather a panel of experts "to survey, interpret and synthesize research findings" on beginning reading and the comprehension of language. Under the direction of professor Richard C. Anderson, the Commission on Reading convened eight leading professors, one first-grade teacher, and one member from the Department of Education to conduct the review. Housed at the Center for the Study of Reading at the University of Illinois, the panel synthesized recent findings from linguistics, cognitive psychology, and child development in a 1985 publication, *Becoming a Nation of Readers*. The report affirmed the value of (1) early language experiences in kindergarten and at home, (2) phonics instruction (in helping children master the alphabetic

principle), and (3) opportunities to read connected text orally and silently. The recommendations emphasized the developmental needs of children as they moved from a basic understanding of the form and function of print to an understanding of letter–sound correspondences and on to independent, fluent reading of books.

In addition to offering recommendations for practitioners, the 1985 NIE report helped reframe the debates about reading instruction in two ways. First, it rejected the dualism between activities designed to foster knowledge of the alphabetic code and opportunities for children to read good literature. New findings from related academic disciplines were shedding light on "the intricacies of the reading process" and "lay at rest once and for all some of the old debates about the role of phonics and comprehension." Second, it raised new questions and encouraged researchers to broaden their study of the reading process and the instructional strategies that facilitated the development of reading comprehension. According to one federal policymaker, the reading research begun by NIE and the publication of the 1985 report "shifted the entire agenda for research and development in that area." It did so by moving the field away from a dominant concern with decoding and early reading instruction to a broader focus on comprehension and language development.

In a second research synthesis, authorized by the federal government and sponsored by the Center for the Study of Reading, Marilyn J. Adams completed a synthesis of the theory and practice of beginning reading. In the 1990 publication *Beginning to Read: Thinking and Learning About Print*, ²⁸ Adams echoed many of Chall's findings and the 1985 NIE report by explaining why phonics instruction facilitated word recognition skills. She also reviewed the growing research literature that undercut the validity of the psycholinguistic theory of reading advanced by Goodman and Smith. Skilled readers used knowledge of letter–sound relationships to process all the graphic information contained in identifying a word. Once lower-level word recognition processes were automated, children could read connected text with speed and focus on comprehending what they read. Thus, the strategy of using context to aid word recognition and the rejection of phonics instruction, as advocated by whole-language theorists, had little support in empirical research by the late 1980s.²⁹

Additional evidence challenging the effectiveness of whole-language practices emerged in a 1989 study published in the peer-reviewed journal *Review of Educational Research*. In this meta-analysis, Steven Stahl and Patricia Miller reviewed the efficacy of whole-language approaches to reading instruction. In addition to informing the scientific literature, the authors hoped to shed

light on the heated debate between proponents of whole-language theories and phonics instruction.³⁰ Stahl and Miller defined whole-language/language-experience methods as having four characteristics: (1) an emphasis on using children's language as a medium for instruction, (2) child-centered lessons, (3) trade books rather than basal texts, and (4) lessons in decoding and phonics only as they arose in the context of reading stories and text.

Their quantitative synthesis revealed that whole-language approaches had some benefits as an instructional approach in kindergarten, but produced inferior results relative to systematic code-emphasis approaches in first grade. To explain this finding, Stahl and Miller suggested that whole language might help children in kindergarten when reading instruction is more romantic and focused on learning the form and function of print. They noted that whole language might be less useful in first grade when children must master the alphabetic code to decode new words. Perhaps most importantly, the findings indicated that whole-language/language-experience methods had the greatest benefit for middle- and upper-class students. Why? According to Stahl and Miller, students from advantaged backgrounds were more likely than low-income students to have learned about the code through exposure to storybooks and language experience at home. Echoing the findings from Chall's earlier research, Stahl and Miller hypothesized that low-income children and poor readers needed explicit instruction in sound-symbol relationships in first grade to become skilled readers.

The Rise and Fall of Whole Language in California

During the late 1980s, research indicating that phonics facilitated efficient decoding skills and that whole language practices lacked evidence of efficacy had little direct influence on state policy and classroom practice. On the contrary, whole-language pedagogy formed the latest conventional wisdom in reading. In a 1990 special issue of *Elementary School Journal* on whole language, P. David Pearson observed that whole language had become a grassroots movement of educators supported by state education officials and professional organizations. Pearson noted that in his 25–year career as an educator,

Never have I witnessed anything like the rapid spread of the whole-language movement. Pick your metaphor—an epidemic, wildfire, manna from heaven—whole language has spread so rapidly throughout North America that it is a fact of life in literacy curriculum and research.³¹

In the same special issue of *Elementary School Journal*, Jerome Harste and Kenneth Goodman argued that one goal of whole-language philosophy and

practice was to empower practitioners. The school curriculum, Harste argued, should not be "left in the hands of those who only rarely come in contact with students." Echoing Harste's ideas about teacher empowerment, Kenneth Goodman asserted that "teachers are not relying on gurus and experts to tell them what to do." The whole-language movement, according to Goodman, was generating a knowledge base "passed from teacher to teacher in person contacts, in teacher support groups, and in local conferences." Rather than following the findings of experimental research published in academic journals, Goodman urged scholars to do research that was useful for teachers, and predicted that "practitioners will move ahead, with or without this support."

By the late 1980s, whole-language theorists had communicated their ideas to decisionmakers in state government in order to change curriculum and instruction. One direct path to changing classroom instruction was to work through state legislatures that had centralized control over textbook adoption policies and the authority to shape the content of the basal texts used in all public school classrooms. According to reading scholar Timothy Shanahan, whole-language advocates were able to persuade the commissioner of education in California, Bill Honig, to adopt new textbooks that de-emphasized skill instruction and phonics skills. According to Shanahan, "[w]hole language-influenced policies translated into a ban on the use of state money to purchase spelling books (whole language proponents opposed spellers),"36 This change in reading curriculum and text represented a radical shift away from traditional basal texts and a move toward child-centered pedagogy. For example, the 1987 California language arts framework supported an integrated language arts curriculum critical of phonics instruction, and advanced the idea that children should construct knowledge on their own, based on their interests.³⁷ It specifically noted that learning English "cannot be limited to a daily list of ten or 15 skill objectives or to the completion of meaningless worksheets."38 Given these recommendations, there was little emphasis on teaching all children to master the 26 letters of the alphabet or the 44 speech sounds that make up the English language. Although some educators embraced the literature-based curriculum and new textbooks, many parents and teachers were alarmed to find that the new materials were too difficult for many children.

The major challenge to whole-language pedagogy eventually came from fourth-grade reading scores on the National Assessment of Educational Progress (NAEP). With the first wave of 1992 data from the newly authorized NAEP administrations at the state level, the federal government supplied the public and policymakers with comparative information on state perfor-

mance. Thus, Californians were alarmed to find that 52 percent of fourth graders read below the basic level on the 1992 administration of NAEP. More bad news followed when the 1994 NAEP scores were released in spring of 1995. Policymakers, parents, and educators learned that the average performance of California's fourth-graders put the state near the bottom relative to other states, and that the decline in scores from 1992 to 1994 was evident among all ethnic and socioeconomic groups. On the 1994 NAEP, 56 percent of fourth graders read below basic, including 46 percent of children from families with college-educated parents.³⁹

The perceived reading crisis was eventually linked to whole-language practice. Survey data indicated that a larger percentage of California teachers employed whole-language practices than their peers in other states. For example, surveys of classroom instruction from the 1992 NAEP indicated that 69 percent of California teachers put a "heavy" emphasis (versus "moderate" or "little or no") on whole language compared to a mean response of 40 percent across other states. Moreover, 87 percent of California teachers indicated heavy reliance on literature-based reading and 52 percent reported little or no reliance on phonics, compared to a mean of 50 percent and 33 percent in other states. Despite the difficulty of drawing firm causal links between instruction and achievement, many policymakers implicated the California language arts framework and poor teaching as the primary cause behind the decline in state reading scores. Despite the difficulty of the primary cause behind the decline in state reading scores.

Since whole language had become the conventional wisdom in reading instruction, any decline (perceived or real) in reading achievement was easily linked to the dominant method of reading instruction. Legislators in California and elsewhere reacted to the educational crisis of low literacy attainment by enacting a flood of phonics bills during the mid-1990s. From 1994 to 1997, 18 states had one or more phonics bills introduced in legislative sessions and California had the largest number of bills introduced during this time period. By 1997, a total of 33 state legislators had passed bills that stressed instruction to improve phonemic awareness or explicit phonics. ⁴² Moreover, the growing number of phonics bills appeared to reflect the public's dissatisfaction with an educational establishment that seemed unwilling to adopt evidence-based practices in the classroom. ⁴³ The reading wars were now being fought in the political arena, as legislators sought to stem the reading crisis by passing laws to govern classroom instruction.

In explaining the political reaction of state legislatures to the reading crisis, Keith Stanovich argued that whole-language theorists had failed to respond to evidence and enact norms of practice rooted in scientific research. In short, whole-language theorists and advocates left the teaching profession

vulnerable to intrusive legislative mandates by failing to police itself. According to Stanovich.

In holding to an irrationally extreme view on the role of phonics in reading education—for failing to acknowledge that some children do not discover the alphabetic principle on their own and need systematic direct instruction in the alphabetic principle, phonological analysis, and alphabetic coding-whole language proponents threaten all of their legitimate accomplishments. Eventually—perhaps not for a great while, but eventually—the weight of empirical evidence will fall on their heads. That direct instruction in alphabetic coding facilitates early reading acquisition is one of the most well established conclusions in all of behavioral science.44

In many ways, Stanovich's criticism of whole language was echoed by California's former commissioner of education. In retrospect, Bill Honig admitted that the 1987 language arts framework and whole-language practices were not based on proven strategies. "It is the curse of all progressives," said Honig, "that we are anti-research and anti-science, and we never seem to grasp how irrational that attitude is. This is probably our deepest failure."45

1993-2000: THE FEDERAL GOVERNMENT TURNS TO EXPERTS TO END THE READING WARS

Political scientist John Kingdon points out that public policy issues and agendas are most likely to capture the attention of legislators when three streams coalesce. 46 In the 1990s, the merging of the problem, political, and policy streams made federal lawmakers eager to convene a panel of experts to prevent reading failure in the early elementary grades.⁴⁷ First, the problem stream flowed from state-level NAEP data in California, which revealed that a majority of fourth-graders could not read at a basic level of performance. Second, the political stream originated from state policymakers' reaction to efforts to win the reading wars by mandating phonics instruction in classrooms where whole-language texts and instruction prevailed. Third, the policy stream emerged from the National Institute of Child Health and Human Development (NICHD), which began to ask and answer timely questions for policymakers and practitioners. In 1993, NICHD encouraged the research community to submit applications for research that addressed the practical question, "Which single treatment/intervention or combination of interventions, provided in which setting or combination of settings, has (have) the most effective impact on well-defined domains of children functioning, for how long, and for what reasons?"48 Under the leadership of G.

Reid Lyon, NICHD advanced its research agenda by funding a programmatic series of experimental and longitudinal studies that showed how appropriate interventions targeted from kindergarten to third grade could reduce failure levels in reading. 49 Converging lines of evidence from basic and applied research began to show that early intervention and direct instruction in phonics could reduce reading failure before third grade.

Given the substantial body of research on the basic processes underlying reading development and the efficacy of different instructional strategies. the National Research Council (NRC) convened leading scholars to synthesize findings from the scientific literature, Published in 1998, Preventing Reading Difficulties in Young Children sought to provide lav audiences and decisionmakers with an integrated picture of how reading skills develop and how to prevent reading failure. According to Catherine Snow, the chair of the NRC report, the consensus about early reading, how it developed, and how instruction facilitated reading ability was "not difficult to reach." 50 The NRC's "core message" to practitioners was that instruction should "integrate attention to the alphabetic principle with attention to the construction of meaning and opportunities to develop fluency."51 In making its final recommendations, the NRC report exhorted scholars and federal lawmakers: "Research toward increasing the efficacy of classroom reading instruction in kindergarten and the primary grades should be the number one funding priority."52 It concluded with 18 additional questions to guide a research agenda on effective primary-grade interventions. Similar to earlier expert panel reports from the 1970s and 1980s, the 1998 NRC report rejected the simplistic dualism between phonics and whole language and raised new questions for fruitful inquiry and research. However, since the panel did not focus its review on questions about the efficacy of different instructional methods. Congress authorized a second panel of experts to conduct an objective review of studies that could provide clear instructional guidance to classroom teachers.

On July 24, 1997, the Senate Committee on Appropriations authorized the director of NICHD to assemble a national panel to synthesize the best research on the effectiveness of different approaches to teaching reading. Senator Arlen Specter stated that the committee was

impressed with the important accomplishments reported from the NICHD research program on reading development and disability, and is eager to have this information brought to the attention of educators, policymakers, and parents.53

The goal of including panelists with diverse professional backgrounds was reinforced in the authorizing statute, which called for a National Reading

Panel of "leading scientists in reading research, representatives of colleges of education, reading teachers, educational administrators, and parents."54

On balance, the 15 panelists were primarily tenured professors in psychology and education and many were leaders in the field of literacy research. 55 Given the unique expertise of each panel member, the NRP eventually formed subgroups that each focused on one of six areas of research: alphabetics (phonemic awareness and phonics instruction), fluency (oral guided reading and independent silent reading), comprehension (vocabulary and text comprehension instruction), teacher education, computer technology, and methodology. By reviewing multiple instructional strategies, the NRP's review implicitly rejected the idea that either phonics or whole-language instruction could produce superior reading achievement. In the words of one of the panelists on the NRP, Congress wanted to settle the "Reading Wars," and put an "end to the inflated rhetoric, partisan lobbying, and uninformed decisionmaking that have been so widespread and so detrimental to the progress of reading instruction in America's schools."56

To make credible causal inferences about the effects of instructional approaches on student outcomes, panelists reviewed only published studies using experimental and quasi-experimental designs.⁵⁷ Given the inclusion criteria, panel member Timothy Shanahan asserted that the report "will be, perhaps, the most thorough and explicit review of these topics ever conducted in reading."58 And by describing in detail the steps that went into the meta-analytic review, the NRP hoped to encourage public scrutiny and review of its procedures and findings.

THE IMPACT OF THE NRP ON RESEARCHERS, POLICYMAKERS, AND PROFESSIONAL ORGANIZATIONS

In 2000, two NRP reports went to press—a 464-page full report with technical details, filled with tables of coding schemes, effect sizes, and p values, and a 33-page summary of the full report. The meta-analysis showed that instruction in phonemic awareness, phonics, and guided oral reading fluency improved children's ability to read words, to read connected text with speed and accuracy, and to comprehend text. Moreover, the report underscored the importance of embedding specific instructional strategies in a comprehensive reading program. For example, although phonics improved word recognition ability, the NRP emphasized 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."59 In addition, the NRP found that providing support and guidance during oral

reading of text helped children improve their ability to read connected text with greater speed, accuracy, and comprehension. However, the NRP cautioned that guided oral reading should be used as part of "an overall reading program, not as stand alone-interventions."60

More broadly, the empirical findings affirmed the vital role that teachers played in improving children's reading skill. The NRP concluded that explicit instruction involving phonemic awareness, phonics, oral guided reading, and comprehension strategies was more effective in improving children's reading skills than student-centered approaches like sustained silent reading, in which children received little or no guidance from teachers in selecting and reading text.61 Thus, the NRP concluded that teacher-directed instruction was essential to improving children's reading, a robust finding that has been documented in over 100 years of education research.⁶²

Since 2000, the NRP's findings have garnered the attention of researchers, policymakers, and practitioners. The findings of the NRP have been widely cited by researchers and subjected to creative re-analyses that have shed light on new questions and influenced federal policy.⁶³ In particular, the NRP's findings on (1) phonemic awareness training, (2) phonics instruction, (3) fluency, (4) comprehension strategies, and (5) vocabulary instruction eventually shaped the Reading First legislation. States and districts must show how federal dollars will support each of the five pillars of scientifically based reading instruction in Reading First schools.64

Professional organizations also helped translate the NRP's findings for its members. The International Reading Association, an organization for reading researchers and practitioners, published a 2002 book titled Evidence-Based Reading Instruction: Putting the National Reading Panel Report into Practice. This edited volume included recent publications from Reading Teacher (an International Reading Association publication with wide circulation to teachers), to guide research-based practice based in each of the five components of instruction reviewed by the NRP, and, according to the editors, "will be a useful tool for educators as they implement practices consistent with scientifically based reading research and the provisions of Reading First."65 For the members of the of the American Federation of Teachers (AFT), the NRP reiterated findings from Louisa C. Moats's 1999 publication, Teaching Reading Is Rocket Science, a lay-friendly publication that summarized research on phonemic awareness, phonics, fluency, vocabulary, and comprehension instruction. Echoing Chall and Adams's earlier works on the value of literature-rich and skill-based instruction, Moats asserted that "teachers need to connect the teaching of skills with the joy of reading, and writing, using read-alouds and the motivating activities popularized by the whole-language movement."66

Because Moats's book had already been disseminated to a large number of teachers, many practitioners were also familiar with the five pillars of scientifically based reading instruction articulated by the NRP report.

Furthermore, the full-length NRP report was condensed into a simpler 33-page summary report for teachers and lay audiences. Critics charged, however, that the summary report misrepresented the findings on the efficacy of phonics instruction. Although the full NRP report provided insufficient data to draw conclusions about the effects of phonics instruction above first grade, the summary indicated that systematic phonics benefited children from kindergarten to sixth grade.⁶⁷ Agreeing that the translation of findings was less than perfect, panel member Timothy Shanahan concurred that the summary was an incomplete and inaccurate summary of the full report. More specifically, he noted that the summary "conveys the idea that good, older readers should be taught phonics, something neither stated nor implied in the report."68 Shanahan added that one remedy to this problem was to make sure that more teachers read the entire full report and enacted evidence-based practices in their classrooms.

Some scholars, however, have challenged the notion that the NRP has promoted good instruction and supported professional autonomy. Richard Allington, a reading professor at the University of Tennessee, charges that the National Reading Panel's findings have been used by proponents of direct instruction and intensive phonics to impose external mandates on teachers. According to Allington, legislative mandates and expert panel reports strip teachers of the autonomy to make curricular and instructional decisions. And if lawmakers and professors have the power to govern curriculum and instruction, Allington wonders whether teachers will continue to feel like autonomous professionals who hold themselves accountable for helping children become independent and skilled readers.⁶⁹

DISCUSSION

"The history of medicine has been written as an epic of progress, but it is also the tale of . . . conflict over the emergence of new hierarchies of power and authority," writes sociologist Paul Starr in his Pulitzer Prize-winning book, The Social Transformation of American Medicine, The Rise of a Sovereign Profession and the Making of a Vast Industry. 70 In many ways, the history of the reading wars might aptly be characterized as an "epic of progress" and a "tale of conflict." The story of progress shows how research findings converged over four decades to form the basis for national policy, most notably in the 2001 Reading First legislation. The story of conflict suggests that research was

also a weak countervailing force in the pendulum swings between whole-language and phonics instruction that took place in state legislatures during the 1980s and 1990s. In the following discussion, my goal is describe how controversies about instruction in beginning reading have been resolved through normal scientific inquiry and why good research alone cannot shape sound instructional policies and practices in reading.

Jeanne Chall's 1967 book Learning to Read: The Great Debate, the National Institute of Education's 1985 report Becoming a Nation of Readers, the 1998 National Reading Council book Preventing Reading Difficulties in Young Children, and the 2000 National Reading Panel report share several qualities of "normal science." According to historian of science Thomas Kuhn, "normal science" builds on "past scientific achievements" and "is sufficiently openended to leave all sorts of problems for the redefined group of practitioners to resolve."71 Major research syntheses of reading traveled down the path of normal science by recognizing the convergent, cumulative, and replicated findings in the scientific research literature. These reports, however, did not immediately or directly impact reading policy and practice. After all, it took the accumulation of three decades of research before a substantive metaanalysis of instruction in phonemic awareness, phonics, fluency, vocabulary, and comprehension could be undertaken in the late 1990s. Nonetheless, the reports showed how substantial agreement in the scientific community was needed before firm recommendations could be made for policymakers and practitioners.

In the conclusion of Learning to Read: The Great Debate, Jeanne Chall asserted that scholarship in reading "should follow the norms of science" by building on the past and raising new questions and hypotheses; a scholar "must try to learn from the work of those who preceded him . . . knowing that neither he nor anyone following him will have the final word."⁷² Chall's observations about the conduct of normal science were realized in the ensuing debates about the merits of phonics instruction. Although Chall found experimental evidence supporting phonics instruction over whole-word reading methods in first grade, Kenneth Goodman challenged these conclusions by speculating that context cues were equally, if not more, important than knowledge of spelling-sound relationships in helping children read new words. Like debates in any scientific field, the novel "psycholinguistic theory of reading" proposed by Goodman sparked the interest of other scholars. According to psychologist Keith Stanovich, "Ken Goodman conducted the well-known 1965 study that focused so many of us in the early 1970s on the study of the effects of context on reading."73 Eventually, Stanovich and others found that reliance on context slowed down word recognition abilities

and was a strategy used by poor readers. Skilled readers, on the other hand, were able to apply knowledge of the alphabetic principle to quickly and automatically read new words. The dispute about the role of context in word reading underscores a truth about scientific progress: No single scholar or individual study dictated the scientific consensus about the processes underlying skillful reading. As Thomas Kuhn points out, the progressive accumulation of research findings produces a discernible "shift in the distribution of professional allegiances" of members in the scientific community.74

That shift in professional allegiances among scholars was captured in consensus panel reports from the 1970s and 1990s, which sought to reject simplistic approaches to reading instruction. In Toward a Literacy Society, a 1975 publication sponsored by the National Institute of Education, Chall argued that neither phonics nor sight-word approaches were sufficient to help children become skilled readers. Instead, she reminded educators and the general public that inflexible approaches "may fail with a child if in the long run it plays down either of these aspects of learning to read. What is important is a proper balance between them."75 A second NIE publication in 1985, Becoming a Nation of Readers, extended Chall's work and synthesized new findings from cognitive psychology and related disciplines. It argued for the need to go beyond word reading and decoding strategies and emphasized the importance of oral language and text comprehension. By broadening its survey of the scientific literature, the report encouraged the scientific community to undertake multi-disciplinary studies of reading and to examine the efficacy of diverse approaches to instruction. In 1998, the National Reading Council publication, Preventing Reading Difficulties in Young Children, recognized convergent findings from diverse scientific disciplines and deepened the foundation on which to base evidence-based reading instruction.

The culmination of nearly three decades of research resulted in the 2000 National Reading Panel report. By raising questions about the efficacy of different instructional approaches and by restricting its review to findings from experiments and quasi-experiments published in peer-reviewed journals, the NRP influenced federal policy and classroom practice. Scientifically based reading instruction needed to focus on word-, sentence-, and text-level outcomes, and claims about the efficacy of different instructional strategies needed the backing of experimental data. Pushing the public and decisionmakers to think beyond the phonics-whole-language dichotomy, the NRP helped reframe definitions about scientifically based research and practice in reading. Moreover, the findings from the National Reading Panel eventually shaped the Reading First legislation in the No Child Left Behind Act, which required eligible Title I schools to adopt scientifically based research practices in five areas of reading instruction: phonemic awareness, phonics, fluency, vocabulary, and comprehension instruction. The five pillars of good reading instruction articulated a new grammar of schooling in education by encouraging practitioners to focus on a broad set of instructional strategies and reading outcomes.

Tracing the evolution of scientific consensus in reading may paint an overly simplistic and teleological version of history—a kind of inevitable and progressive accumulation of research toward the ultimate goal of improving instruction and achievement. However, the reading wars are also a tale of conflict and of pendulum swings between externally mandated wholelanguage practices and phonics legislation during the 1980s and 1990s. The story of conflict offers some reasons why research alone cannot protect educators from unproven theories and policies, whether originating in the minds of an academic researcher or the actions of a state education official. In particular, I elaborate on three characteristics of scholarly research and the scientific community that help explain why research does not immediately influence and shape education policies.

First, the results of normal scientific inquiry are usually reported in peer-reviewed journal articles, slowing down the dissemination of scientific research findings to decisionmakers and lay audiences. From the 1960s to 1980s, scientists began to undertake basic research on the processes underlying skillful reading in laboratories, and applied research on the efficacy of different instructional approaches in classroom settings. However, it took several decades for researchers to highlight convergent findings and make recommendations for teachers. 76 Louisa C. Moats has observed, "there is always a long delay between developments in academic research disciplines and their incorporation into teaching practice."77 Research often takes several decades to bear fruit, but decisionmakers cannot wait for decades to help struggling readers. Consequently, the demands facing a state education official, superintendent, or teacher create pressures for immediate action and quick solutions. Research that is unavailable for decades cannot inform decisionmaking today.

Speeding up the process by which scientific controversies are resolved may equip practitioners with more relevant and timely information. Adversarial collaboration represents a recent effort among scientists to accelerate the process for resolving controversies. Expounded in the journal Psychological Science, the editors expressed hope that adversarial collaboration would become a more widely used protocol for adjudicating disputes between scholars and disseminating findings quickly to avoid ongoing controversy.⁷⁸ The procedure requires adversaries in a scholarly debate to agree on basic design

issues and research questions before they conduct the study. In addition, it requires antagonists to collaborate on a prospective study and agree on an arbiter who imposes the rules of engagement over the entire process. The arbiter helps adversaries decide on the design of the experiment, controls the data, determines the final venue for publication, and can even declare in the final publication if an uncooperative participant failed to comply with the agreed-upon protocol. In other words, adversarial collaboration represents a potentially valuable and under-utilized tool for mediating conflict in scientific debates. In the future, it might help to adjudicate debates in reading and the many education policy controversies outlined in this volume.⁷⁹ Ideally, encouraging adversaries to collaborate on prospective studies would accelerate the resolution of conflict in the research community and provide the kind of scientific consensus that informs good practice in schools.

Second, normal science depends on the validation of research findings by a community of experts who are expected to remain objective participants in democratic debates. Scientists are asked to educate, not advocate. "One of the strongest, if still unwritten, rules of scientific life is the prohibition of appeals to heads of state or to the populace at large in matters scientific," observes Thomas Kuhn.⁸⁰ These strong professional norms create disincentives for scholars to jump into the policy area and advocate for specific policies or educational curricula. Therefore, the mere existence of good research is no guarantee that such knowledge will be communicated to policymakers. Nonetheless, the critical perspective of scientists performs a valuable function in a democracy where public policies are crafted by politicians in local school boards, state legislatures, and Congress. Scientists do not have power or authority to mandate phonics instruction, to adopt literature-based basal texts, or to define scientifically based reading instruction in federal statutes. They can, however, encourage legislators to evaluate untested policies before they are brought to scale.

For example, the critical perspective of a social scientist played a vital role in causing Tennessee legislators to require an evaluation of class size reduction in the mid-1980s. Although state lawmakers wanted to enact class size reductions to improve student achievement in the early grades, Steven Cobb, a sociologist by training, encouraged a randomized experiment to evaluate the efficacy of small class sizes on student learning.81 Tennessee lawmakers eventually passed legislation to undertake a statewide experiment called Project STAR (Student Teacher Achievement Ratio), which has been hailed by scholars as one of the most influential studies in education.⁸² "The role of social science research," writes Daniel Patrick Moynihan, "lies not in the formulation social policy, but in the measurement of its results."83 When sociologists like Steven Cobb and Daniel Moynihan enter the political arena, they often contribute to democratic debate by encouraging critical evaluation of unproven and costly policies. Had social scientists shared findings from Stahl and Miller's meta-analysis of whole language with state legislators, perhaps California's state officials would have called for a smaller pilot study or an evaluation of whole-language practices before it was implemented in all districts and schools.

Third, the federal government has frequently turned to scientists, not teachers, to determine evidence-based practice. The assumption, of course, is that scientists have the tools and knowledge to understand research and can therefore establish norms of practice that buffer the teaching profession from fads, ideology, and political intrusion. To practitioners, however, experts can be viewed as novices who know little about teaching in the public schools, and effect sizes from a meta-analysis can be viewed as irrelevant tools for addressing the range of skills and performance in a classroom of 30 first-graders. Indeed, Joanne Yatvin, the lone practitioner on the National Reading Panel, wondered how professional standards could be determined primarily by scholars who do not teach children how to read. In filing a Minority View in the appendix of the NRP, Yatvin argued that panelists failed to subject their results to the scrutiny of teachers. "Outside teacher reviewers," Yatvin argued, "should have been brought in to critique the panel's conclusion, just as outside scientists were to critique its processes."84

Without being represented on these expert panels, teachers and their allies have frequently asserted that external mandates by federal and state lawmakers and consensus reports by university researchers undercut the professional autonomy of K-12 teachers. Professionally eclectic expert panels in reading-perhaps even an equal number of teachers and professors-might address these criticisms by giving voice to teachers. For example, by including an equal number of teachers and nonteachers on the United Kingdom's National Literacy Task Force, political leaders encouraged recommendations for improving reading instruction that integrated the practical knowledge of teachers and findings from researchers.85 Recent efforts to bridge the gap between researchers and practitioners provide hopeful signs that collaborative efforts may deepen the legitimacy of scientific evidence among teachers and encourage researchers to pursue answers to relevant, practical questions. For example, the International Reading Association (IRA) and the National Institute of Child Health and Human Development (NICHD) jointly sponsored a research seminar on teaching English-language learners. According to Peggy McCardle, the chief of the Child Development and Behavior Branch for NICHD, the long-term goal of these collaborative meetings is to "get

researchers talking both to each other and to practitioners."⁸⁶ In addition to fostering dialogue, inviting teachers and researchers to make policy may empower teachers to shape the norms governing their profession.

CONCLUSION

By design, normal science proceeds slowly and convergent findings take decades to evolve. By tradition, scientists must embrace neutrality in public policy debates and avoid partisanship in controversies about reading instruction. And by necessity, government has usually turned to university professors to translate scientific research and technical findings in academic journals for classroom teachers and lay audiences. The defining characteristics of normal science are in many respects virtues of the scientific enterprise. Waiting 30 years for scientists to conduct enough studies to be included in the National Reading Panel's meta-analyses seems worthwhile if these research findings are helping to improve the quality of teaching and learning in classrooms. Recent efforts to speed up the resolution of scientific controversy, to encourage communication between social scientists and policymakers, and to forge collaborations between research and practitioner communities may help resolve conflict in reading and other areas of education policy and practice. In the long-run, it is unclear whether any of these initiatives will create an enduring peace in the reading wars.

Perhaps the surest path to protecting reading policy and practice from radical pendulum swings, fads, and ideology is to create a sovereign profession. Ultimately, teachers must be involved in establishing and regulating professional norms. Sociologist Paul Starr asserts that the legitimacy of professional authority and competence rests on "three distinctive claims: first, that knowledge and competence of the professional have been validated by a community of his or her peers; second, that this consensually validated knowledge and competence rest on rational, scientific grounds; and third, that the professional's judgment and advice are oriented toward a set of substantive values, such as health."⁸⁷

Few would dispute that the teaching profession is oriented toward a substantive and valuable goal—the education and cognitive development of young children. Today, it would also be noncontroversial to suggest that substantial progress in reading research has built a strong empirical foundation for improving reading instruction. Yet the first claim of professional authority—the validation of professional competence by a community of peers—remains an elusive goal in American education. Among the professions, teachers remain in the unenviable position of lacking the power and authority to

insulate good practice from the misguided theories of academic researchers or the faddish policies of political leaders.

In the future, will a community of practitioners validate the knowledge and ability of their colleagues' to instruct children how to read? Will teachers belong to a sovereign profession that compels its members to meet norms of excellence agreed upon by a community of peers, applies scientific research in shaping professional standards, and serves its clients well? Or will teaching remain a partial profession where professors and lawmakers possess the primary authority to mandate policy and shape practice? Empowering teachers to establish professional norms rooted in scientific research may help create a sovereign profession. Ultimately, teachers must have access to truth and power if they are to establish professional norms that support their efforts to help children become skilled readers and active participants in our democracy.