

## 24 Reading Policy in the Era of Accountability

*Cynthia E. Coburn, P. David Pearson, and Sarah Woulfin*

University of California, Berkeley

The last two decades have witnessed a dramatic upsurge in policy making related to reading instruction. Several high profile policy initiatives place reading instruction squarely at the center of reform policy; moreover, those initiatives employ new policy mechanisms that reach down into the classroom in unprecedented ways. In the time since publication of the third edition of the *Handbook of Reading Research*, Congress passed No Child Left Behind, spreading high stakes accountability policy to states and districts across the nation. Since reading achievement sits at the core of many state accountability policies, NCLB has dramatically raised the stakes for student performance on standardized tests in reading, placing new pressure on teachers, schools, and districts to improve reading achievement.

In addition, many states and districts have responded to these accountability demands with their own policy making. For example, districts across the country have adopted new reading textbooks districtwide, linking them with pacing guides, increased professional development, and progress monitoring assessments. Finally, this time period has witnessed two federal initiatives specifically devoted to improving reading instruction: Reading Excellence Act and, more recently, Reading First. Reading First is distinctive in the history of educational policy making for the degree to which the federal government and states have specified appropriate instructional practice, focused on fidelity to curricular materials, and introduced extensive monitoring of teacher practice.

The increased policy attention on reading instruction and the use of new and aggressive policy instruments is bringing policy into classrooms as never before. Now, more than ever, the study of instructional policy is crucial for understanding the nature of reading instruction in our nation's public schools. To that end, this chapter provides a comprehensive review of empirical research on reading policy since publication of the third edition of the *Handbook of Reading Research*. After a brief review of recent trends in state and federal policy making in reading, we review studies related to elementary reading instruction in three parts: the dynamics of policy making in reading, the relationship between reading policy and teachers' classroom practice, and, ultimately, the impact of these policy initiatives on student achievement. In so doing, we trace the pathways from policy making through implementation and, finally, policy outcomes. Throughout, we paint a portrait of the state of research on reading policy as growing, but incomplete. We highlight key findings and point out areas that are ripe for future inquiry.

### OVERVIEW OF READING POLICY

Policy makers have engaged in unprecedented efforts to use policy as a lever to improve teaching and learning in reading instruction in recent years. Throughout the 1990s, the country was in the midst of the evolving standards movement. Encouraged by federal Goals 2000 funds, state after state enacted new state standards in different content areas including reading and language

arts. Based on the logic of systemic reform (Smith & O'Day, 1990), these standards tended to put forth ambitious visions of teaching and learning, which were then linked with state assessments to monitor progress and professional development to build teacher capacity. The basic idea was for states to put forth a vision of valued learning outcomes and use standardized tests to measure the degree to which schools met those outcomes, at the same time allowing local schools and districts the flexibility in deciding *how* to meet those outcomes (Cross, Riley, & Sanders, 2004). By 1995–1996, 48 states and the District of Columbia had developed academic standards and 42 states had implemented or were developing assessments aligned with academic standards (Gandal, 1996). For example, Michigan enacted a series of legislative mandates in the late 1980s and early 1990s that leveraged Goals 2000 funds to develop an academic core curriculum in key subject areas. The result was the Michigan English Language Arts Framework (MELAF), which developed integrated standards in English Language Arts. Those standards were intended to guide district efforts to develop a core curriculum as required by law (Birdyshaw, Pesko, Wixson, & Yochum, 2002; Borman & Cusick, 2002). The state subsequently used MELAF to guide revision of the state assessment—the Michigan Educational Assessment Program (MEAP)—so that state assessments were aligned with state standards. Finally, the state embarked on a program of professional development to local districts to build district and school capacity to enact the vision of ambitious teaching and learning embodied by the standards (Dutro, Fisk, Koch, Roop, & Wixson, 2002; Jennings, 1996).

At the same time that states were engaged in adopting standards in reading and other content areas, low reading scores on the 1992 and 1994 National Assessment of Educational Progress led to widespread perception of a crisis in reading instruction by the mid-1990s (Colvin, 1995). States responded by enacting additional legislation that focused specifically on early reading instruction. Most of these laws provided funds to support professional development for teachers and the purchase of instructional materials that focused on phonics and phonemic awareness. For example, in 1995, the state legislature in California adopted AB 170, which required that basic instructional materials in math and reading in grades 1–8 emphasized basic skills including phonics and spelling. Later, they adopted two additional bills (AB 3482 in 1996 and AB 1086 in 1997) that appropriated \$89.4 million for professional development emphasizing phonics and phonemic awareness for K–3 and later 4–8 teachers throughout the state (California State Board of Education, 1999). By 1999, 36 states had bills that were passed or pending that promoted the use of phonics-based materials or provided professional development for teachers in instructional approaches to phonics or phonemic awareness (Pearson, 2004).

By the late 1990s, the federal government got in the action and began passing legislation specifically targeting reading instruction as well. In 1998, congress passed the Reading Excellence Act (P.L. 105-277) with the stated purpose of improving students' reading skills and teachers' instructional practice by using methods based in scientifically-based reading research (Edmondson, 2005; Mesmer & Karchmer, 2003). This legislation created a program of competitive state grants to provide staff development and tutorial assistance to low-performing and high-poverty districts. All professional development approaches and materials for children paid for under this grant program had to be supported by research that was considered "scientifically based." In this case, "scientifically based" was defined as employing systematic methods, using rigorous data analysis, valid and reliable measures, and being published in peer-reviewed publications. The legislation was significant because it was the first time that the federal government legislated a definition of reading instruction. It also marked the beginning of the press for "scientifically-based" research in federal legislation in education (Edmondson, 2005; Eisenhardt & Towne, 2003).<sup>1</sup> Like standards-based reform, this approach to policy making at the federal and state level tended to place its bets on impacting teachers' classroom practice and student learning by focusing on the materials that teachers used and building teacher capacity through professional development. However, these policies also tended to specify instructional practices with much

greater precision than state standards and, in some cases, states linked funds with the use of specific curriculum materials, often those certified to be based on scientific research.

As the 1990s drew to a close, policy makers at the state, federal, and even district level began to shift toward a new policy strategy to improve reading instruction: test-based accountability. While state standards remained in place, this shift placed increased emphasis on testing by linking performance on standardized tests to a program of sanctions and rewards (Hamilton, Stecher, & Klein, 2002). Reading instruction was often at the center of these policy initiatives because most states include test scores in reading as a central part of their accountability formulae.

Several states led the way with high stakes accountability policies focused on reading instruction. For example, in 1993, the state legislature in Texas adopted their state accountability system. This system included the adoption of a statewide curriculum known as Texas Essential Knowledge and Skills (TEKS), a statewide assessment system, and the development of an extensive data system to monitor school and student progress (Haney, 2000; Vazquez Heilig, & Darling-Hammond, 2008). The accountability system was then updated in 1999 to better align TEKS with the state assessment, expand testing to all students in grades 3–8, and link passing the state test in reading and mathematics with promotion in grades 3, 5, and 8 (Texas Education Agency, 2002). A central feature of the accountability system in Texas was school performance ranking. All schools were assigned the rank of exemplary, recognized, acceptable, or unacceptable based on their test scores in reading and mathematics in elementary and middle school, dropout rates, and student attendance.<sup>2</sup> In a move that foreshadowed No Child Left Behind, student performance data were disaggregated by ethnicity and socioeconomic status, and school rankings were based on student performance in all categories. Schools were awarded cash awards for high performance and were subject to sanctions, including the possibility of school closure, if they missed performance targets for two years in a row (Haney, 2000).<sup>3</sup> Although each a bit different in design, Massachusetts, Kentucky, and Maryland were other early adopters of high stakes accountability policies as were some local school districts such as Chicago and Philadelphia (Mintrop & Trujillo, 2005).

High stakes accountability subsequently became the predominant policy strategy across the country with the passage No Child Left Behind (P.L. 107-110), which was signed into law in 2002 after receiving widespread bipartisan support in Congress. This sweeping federal legislation tied receipt of federal Title 1 funds to the development of a single statewide accountability system with annual tests in reading and mathematics (and, later, science) for students in grades 3–8. States were also required to set improvement targets such that all students reach proficiency by 2014. And, they were required to create a series of increasingly serious sanctions such as state takeover if schools failed to make improvement targets for all subgroups of children (Fuhrman, 2004; Stecher, Hamilton, & Gonzalez, 2003). This legislation brought high stakes accountability—heretofore present only in select states and districts—to all states across the country.

A final major policy strategy implicating reading instruction also emerged in the late 1990s and early part of the 21st century: mandated curricula. This policy strategy appears to have developed as local school districts in different regions of the country began to see mandated curricula as a strategy to improve reading achievement. Districts across the country not only began to mandate a single curriculum for the entire district, they began to link the curricula to pacing guides, progress monitoring assessments, and monitoring to ensure fidelity to the instructional approaches embodied in the chosen curriculum (Sloan, 2006; Snipes, Doolittle, & Herlihy, 2002; Achinstein & Ogawa, 2006).

This strategy subsequently gained a much higher profile when it was embedded in the federal Reading First policy. Reading First, part of NCLB (Title 1, Part B, Subpart 1), was a competitive state grant program intended to assist low-income, low-performing schools in raising student reading achievement (USED, Reading First Implementation Evaluation, 2008). To receive

funds, states had to develop plans for increasing teachers' use of scientifically-based instructional approaches through the adoption of scientifically-based curricular materials.<sup>4</sup> Drawing on the influential meta-study from the congressionally-mandated National Reading Panel (NRP), Reading First also focused attention on what has come to be called the "big five": phonemic awareness, phonics, fluency (instantiated as either guided reading instruction or independent reading), comprehension, and vocabulary.<sup>5</sup> Finally, Reading First emphasized diagnosis and prevention of early reading difficulties through the use of valid and reliable assessments, interventions for struggling students, and ongoing monitoring of struggling students (Gamse, Jacob, Horst, Boulay, & Unlu, 2008; United States Department of Education, 2002; United States Department of Education, 2004).

Like the Reading Excellence Act, Reading First provided funds for the use of "scientifically-based" materials and professional development to ensure that teachers used them as intended. But Reading First also intensified the policy intervention by linking the curricular materials to the use of progress monitoring assessment (for example, the DIBELS assessments, which were adopted by many states), providing stronger guidance about how curricular materials should be used, and instituting the use of school-based coaching to provide on-site professional development. Many states also stipulated that teachers in Reading First schools use the adopted curriculum and interventions with fidelity and put procedures in place to monitor teachers' instructional practice to ensure fidelity to implementation. For example, Massachusetts' Reading First program proposed a comprehensive instructional program, including state-selected core, supplemental, and intensive intervention programs, aligned with reading research as represented in the NRP report and the state's English Language Arts framework (Massachusetts Department of Education, 2005a, p. 3). To monitor implementation at the state, district, and school levels, the state department of education relied on their consistent assessment system and implementation evaluations. The state conducted school site visits, in part to monitor implementation of the use of core reading program, materials, and assessments (Massachusetts Department of Education, 2005b, p. 17). Reading First also unfolded in the context of high stakes accountability provided by NCLB, so the underperforming schools targeted by Reading First were often experiencing the increased sanctions associated with NCLB as well.

It is important to note that with Reading First, research played a new role in impacting policy, curriculum, and practice. In the past, research had 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 case of Reading First, scientific research was the driver of the system and the basis on which the standards for the primary policy levers—curriculum, assessment, and professional development—were established. This was not the benign, "sowing the seeds of knowledge" logic of marketplace of ideas. It was reform by mandate and monitoring. Above all, the enactment of Reading First in NCLB meant that research had become a full participant in the policy arena.

Three final things are worth noting about the policy-making activity during this time period. First, it was extensive. The last two decades has witnessed unprecedented policy attention to reading instruction as states and the federal government have become increasingly and actively involved in crafting policies and programs intended to improve student achievement in reading instruction. Second, it was layered. Each wave of policy activity, rather than supplanting earlier policy initiatives, tended to layer new policies on top of pre-existing ones. For example, the adoption of state standards during the standards movement in the late 1980s and early 1990s became linked to high stakes assessment at the dawn of the accountability movement in the mid 1990s (Fuhrman, 2004; Hamilton *et al.*, 2002). Many districts responded to the accountability movement by mandating districtwide curriculum adoptions, which they then monitored with benchmark assessments and attention to test score performance (Achinstein & Ogawa, 2006). Reading First intensified this trend for those schools that received Reading First monies. As

a result, schools tended to experience the cumulation and intensification of policy initiatives. Finally, the policy making was serious. The past decade and a half witnessed the use of new and more aggressive policy instruments. Over the course of different policy movements, guidance to schools and teachers became increasingly specified as it moved from the logic of standards (putting forth broad goals and assessing outcomes, but leaving flexibility to local sites in enacting those goals) to mandated curriculum linked with assessments, pacing guides, and close monitoring for fidelity. At the same time, stakes for student performance have increased over time with the advent of test-based accountability linked with sanctions and rewards. These waves of policy have reached into classrooms to change the core activities of reading instruction.

## LITERATURE REVIEW

In the rest of this chapter, we review studies that investigate the causes, processes, and consequences of these policy initiatives. In bounding our review, we focused on studies concerned with reading policy that implicated instruction in grades K through 5 and that were published from 1999 until the present. To be included in the review, articles needed to (a) be empirical (rather than essays or opinion or advocacy pieces), (b) have been published in peer-reviewed journals, and (c) have met standards for scholarly articles (for example, we excluded articles that did not have a methodology section). We made one exception to these criteria. Because of the proximity of this review to passage, enactment, and demise of Reading First, we found very few articles published in peer-reviewed journals that investigated Reading First. Given the importance of this policy initiative, we decided to include federally-funded evaluation studies in five states with some of the largest Reading First grants (California, Texas, Florida, Illinois, and Pennsylvania)<sup>6</sup> as well as select cross-state evaluations of Reading First published by research firms that secured contracts to conduct the evaluations. To locate peer-reviewed journal articles, we searched the indexes for all of the major journals of reading instruction, policy, leadership, and general education from 1999 until the present.<sup>7</sup> We also reviewed reference lists in articles we located to point us toward other articles that might have missed. All told, we identified 121 studies of reading policy during this time period and included 87 articles that met our criteria in this review.

In the sections that follow, we review this literature in three parts. First, we review research on the dynamics of policy making in reading instruction. This small body of studies investigates the process of policy making, seeking to uncover how some ideas become embedded in policies and not others and the role of key actors in this process. Given the attention to scientifically-based research in recent legislation, we pay careful attention to studies that investigate the role of research in policy-making activity. Second, we review studies that investigate the process of policy implementation. Here, we review studies that investigate how teachers' respond to policy initiatives in their classroom and the individual, organizational, and policy factors that influence when and how teachers respond in particular ways. Finally, we review studies that investigate the impact of policy on student achievement, with a special focus on the impact of Reading First.

## READING INSTRUCTION AND THE POLICY-MAKING PROCESS

As reading instruction has become a high profile part of education policy making at both the state and federal level, it has become of increasing interest to reading researchers. In the time since the last *Handbook* edition, reading researchers and others have begun to turn their attention to the dynamics of policy making in reading, perhaps for the first time. Much of the writing about policy making by reading researchers focused on analyzing the degree to which reading



legislation is rooted in reading research. This is not surprising given that claims that particular practices supported by legislation are “research-based” and arguments about which research is “scientifically-based” have become a central feature of contemporary reading policy debates (Eisenhardt & Towne, 2003; Pearson, 2004). Beyond analysis of the degree to which legislation is research-based, a small group of scholars have also investigated the process by which particular ideas about reading instruction became a part of policy initiatives, asking: What is the process by which some ideas and not others become prominent in the policy arena and become embedded in legislation? Here, we review both of these lines of research.

Most of the literature on policy making in reading has focused on the role of reading research in the policy-making process. Some of this work has analyzed actual pieces of legislation—Reading First, the Reading Excellence Act, or state reading frameworks—to determine the degree to which they are rooted in research (Camilli *et al.*, 2003; Camilli *et al.*, 2006; Pearson, 2004; Pressley & Fingeret, 2007). Other researchers analyze documents that have played prominent roles in the policy-making process, like the National Reading Panel Report (NICHD, 2000) or Bonnie Grossen’s “white paper” (1997), which appears to have played an important role in the formation of the California reading policy in the late 1990s (Allington, 1999; Allington & Woodside-Jiron, 1999; Dressman, 1999; Pressley & Fingeret, 2007; Snow, 2000; Taylor, Anderson, Au, & Raphael, 2000). The general strategy of this genre is to examine the claims about best practices in reading put forth in these documents and legislation and analyze the degree to which they are supported by the existing research on reading. Although there is definitely disagreement about the degree to which various documents and legislation are rooted in research (see for example, exchange between Snow, 2000 and Gee, 1999), these studies find a pattern of overstatement of the strength of research findings (Allington & Woodside-Jiron, 1999, Camilli *et al.*, 2003; Camilli *et al.*, 2006; Pressley & Fingeret, 2007; Taylor *et al.*, 2000) and an extrapolation of programmatic solutions from research on student learning, especially in legislation (Allington & Woodside-Jiron, 1999; Pearson, 2004). However, these articles do not analyze the actual policy processes involved. That is, they do not investigate how the particular vision of reading instruction came to be so prominent in state and federal legislation. They do not investigate how particular documents were used in the policy-making process or why these documents and not others came to be influential.

Fortunately, a small body of research has emerged during this time period that sought to address precisely these questions. First, several articles by Cecil Miskel and his colleagues sought to understand the dynamics of federal reading policy during the time leading up to passage of Reading First. Using social network analysis, document analysis, and interviews with key policy informants, this line of work investigates the dynamics of reading policy “issue networks.” Issue networks are webs of linkages between people who are involved in and knowledgeable about a particular policy arena. Prior research outside of education has provided evidence that issue networks play a crucial role in policy making because they are the key mechanism through which those inside the policy system (legislatures, congressional staffers, White House staff, etc.) connect with those outside the system (advocacy organizations, researchers, professional organizations, etc.). Through their formal and informal relationships, those in issue networks influence others to support their point of view and mobilize to support particular policy solutions and not others. It is in the context of issue networks that public policies are debated, refined, and negotiated (Hecklo, 1978; Kingdon, 1984; McFarland, 1992).

Miskel and his colleagues (McDaniel, Sims, & Miskel 2001; Miskel & Song, 2004; Song & Miskel, 2005) found that in the late 1990s, the national issues network for reading instruction expanded greatly. They identified 131 organizations that were actively involved in shaping reading policy at the national level. These organizations included traditional reading policy actors such as reading professional organizations and the national teachers unions. But they also found that a host of new actors—including representatives of the business community, the medical commu-

nity, the special education/disabilities community, and advocacy groups representing the needs of poor children and children of color—had become active players. However, of these 131, Miskel and his colleagues identify a smaller number—18 organizations and 5 individuals—who those in the reading issue network agree are most influential in the policy debates, including the National Institute for Child Health and Human Development (NICHD), American Federation of Teachers (AFT), International Reading Association (IRA), and the National Education Association (NEA) and such individuals as Reid Lyon from NICHD, Congressman Bill Goodling (who had previously sponsored the Reading Excellence Act), Bob Sweet who was a congressional staffer, Louisa Moats from NICHD, and Marilyn Adams, a prominent reading researcher.<sup>8</sup> Interview data suggests that those who were particularly influential were skilled at disseminating research synthesis that promoted their point of view, including some of the documents that were analyzed in the research reviewed above (e.g., the National Reading Panel Report and the Grossen piece). They also collaborated with one another to bring their message to the fore, had both formal and informal contacts with policy makers in key decision making roles and, perhaps most important, were perceived by policy makers as objective (McDaniel et al., 2001).

But while reading policy activity in the second half of the 1990s was characterized by broad participation, the development of the Reading First legislation in late 2001 was a different story. Miskel and Song (2004) used social network analysis and interviews to map out the issues network for the 18 organizations and 5 influential policy actors identified in the first study. They found that the issue network for Reading First was characterized by a core-periphery structure. At the core were a small number of government actors, especially those in President George W. Bush's domestic policy office who were responsible for drafting the Reading First legislation and moving it through Congress. These core actors were connected to "idea champions" who were non-governmental actors such as Reid Lyon. But, overall, there were remarkably few actors involved in actually drafting and debating the Reading First legislation. This configuration of an issue network is unusual in the literature. However, Miskel and Song (2004) argue that it was made possible because of the consensus that was developed among influentials in the issue network during passage of the Reading Excellence Act in 1998. In the absence of organized opposition or conflict, the insider approach to policy making flourished.

In addition to the work at the federal level, several researchers have also examined the process of policy making at the state level (Allington, 1999; Coburn, 2005a; Cusick & Borman, 2002; Song & Miskel, 2005). Although not all studies use the concept of issue networks, the findings are remarkably consistent with Miskel and his colleagues' analysis at the federal level. These studies paint a portrait of policy making at the state level as occurring in the interaction between government actors and a network of professional organization, advocacy organizations, and university actors. The configuration of these actors and the nature of the positions that they promote influence the focus and content of policy.

State level comparisons suggest that while the size of issue networks and the configuration of key actors varied by state, government actors were generally much more influential (i.e., had greater measures of centrality and prestige) than professional organizations and other interest groups across states (Song & Miskel, 2005). Interview data suggests that limited influence of professional organizations at the state level is due to the fact that many state-level teacher organizations are not focused on the specific content area of reading. Thus, while they are active in state education policy generally, they are less active in subject-matter specific policy making. Furthermore, respondents suggest that many of these organizations tend to focus on implementation issues rather than policy making, limiting their influence in the policy-making process.

There is some disagreement about the degree to which professional reading organizations play an influential role at the state level. While Song and Miskel (2005) provide evidence that teacher organizations had low centrality and prestige in reading issues networks in the eight states in their study including Michigan, Cusick and Borman (2002) paint a different portrait

of the role of professional organizations in that state. They use a longitudinal, ethnographic study of policy making in Michigan to provide evidence of the widespread and central role of a network of professional organizations and teacher networks in creating and promoting the state reading framework, although the influence was filtered through bureaucratic authority of the state department of education, which, in turn, mediated political pressure from the state board of education. One explanation for this discrepancy is the fact that the two studies were done at two different historical time periods. Cusick and Borman's study was conducted from 1993 to 1996 and followed the development of the Michigan English Language Arts Framework. This framework, which promoted constructivist approaches to reading and writing instruction, grew out of what Cusick and Borman characterize as a social movement involving reading professionals and university researchers that had been developing consensus on this approach to instruction for years. These outside organizations were able to mobilize quickly when the opportunity to write the Frameworks arose. In contrast, Song and Miskel (2005) studied policy making in Michigan during the early 2000s, when a different approach to reading instruction was on the state policy agenda and many states experienced a shift in the configuration of actors who were influential in promoting these policy approaches (Allington, 1999; Coburn, 2005a).

Indeed, Coburn's study of shifts in California reading policy from 1983 to 1999 provides evidence that different policy eras have distinctly different configurations of actors. For example, she documents how policy shifts in the late 1980s brought together a grassroots teacher movement, professional organizations, and key university researchers with top-down policy initiative led by then State Superintendent Bill Honig to create reading policy promoting literature based instruction. In contrast, in the mid 1990s, efforts to promote greater attention to phonics and phonemic awareness in state policy were initially promoted by a small group of advocates, policy makers, and researchers (many from out of state). Many of the extensive network of actors who were involved in earlier policy initiatives in California were excluded from legislative hearings and task forces in this new period of policy making, and few of the positions they advocated for were considered as part of the policy-making process.<sup>9</sup> Thus, there were different people from different professional and disciplinary backgrounds who were influential in state policy making at different times, accounting for the sharp shifts in California state reading policy in a relatively short period of time.

Taken together, these studies present a first look at the dynamics of policy making in reading during an especially active and fertile time in reading history. But, there is still much to learn. Given the importance of issue networks, we need more studies that help us understand how they are formed and transformed. Existing studies document that they exist, that they are influential, and that they change over time. But we have little information about how and why they exist in the form that they do and what forces cause them to change over time. Similarly, we know that there is often a rather loose relationship between research on reading and the policy-making process, yet some bodies of research do play a role in reading policy. More studies are needed to understand when and under what circumstances research findings move through networks and into the hands of government actors who are crafting policy positions. It is only by understanding the processes by which research moves in and through the system and into policy and how its meaning is transformed along the way that reading researchers and others can understand how they can act to form more productive relationships between researchers and policy makers in the service of improving reading instruction for children.

## POLICY IMPLEMENTATION

While there is only limited research on the policy-making process in reading instruction, there is an ever-growing body of research that seeks to understand how reading policy, once enacted,



moves down and through the policy system and into schools. Most of this scholarship has addressed one of two questions. First, scholars have asked: How are teachers responding to these new reading initiatives? In particular, how are new policy initiatives being implemented in the classroom? And second, some scholars have also asked: What are the factors that influence the implementation process, shaping the degree to which teachers take up new policy ideas in substantive ways? What factors account for teachers' ability and inclination to change their practice in response to policy demands? These questions are critical because they open up the black box between policy enactment and policy outcomes. The answers to these questions provide insight into when and under what conditions some policy initiatives lead to increased student learning and others fail to reach the classroom at all. Thus, studies of the policy implementation process have the potential to provide insight into key points of leverage for supporting and sustaining change in reading instruction over time.

In the sections that follow, we begin by reviewing the findings on how teachers responded to reading policy, with particular attention to the degree and nature of changes in their classroom instruction. We then move on to investigate the factors that shaped how teachers responded, considering individual factors, features of the social context, and features of the policies themselves.

### ***Teachers' Response to Reading Policy***

Research during the last decade has investigated teacher responses to a range of policy initiatives as reading policy has shifted rapidly throughout the late 1990s and the early part of the 21st century. In this section, we review the evidence of teachers' response to each kind of policy initiative in turn. We move chronologically, starting where the previous edition of the *Handbook* chapter left off with a review of teacher responses to new state standards, then moving on to review teachers' response to high stakes assessment, and finally to a look at mandated approaches to curriculum implementation that are characteristic of both Reading First and other recent state and district policy initiatives.<sup>10</sup>

#### ***Standards***

Only a small number of studies during the period of the review investigated teachers' response to the introduction of new state standards in reading. These studies tended to converge around a single important finding: in implementing state standards, teachers tended to focus on surface features rather than higher level learning outcomes and deeper pedagogical changes promoted by the standards, resulting in superficial implementation (Coburn, 2004, 2005a; McGill-Franzen, Ward, Goatley, Machado, 2002; Ogawa, Sandholtz, J. H., & Scribner, 2004; Spillane, 2000). For example, Coburn (2004) studied three California teachers' response to multiple policy messages about reading, including those related to the new state standards. Based on her analysis of teacher responses to 223 policy messages, she developed a typology of teacher responses that ranged from rejection to symbolic response (changing appearances without changing instruction), parallel programs (layering new approaches on top of existing ones), assimilation (transforming new approaches so that they resemble prior practice), and accommodation (restructuring prior practice). She found that the most common response to policy messages by far was assimilation. That is, teachers—even those who were supportive of the reform efforts—came to understand new policy through the lens of their pre-existing practices and beliefs. In so doing, they tended to focus on changing surface aspects of their instruction such as the materials they use or classroom organization while leaving their underlying pedagogical approaches intact. Teachers responded to 49% of all policy messages by assimilating them into their existing instructional approaches compared to 9% of policy messages by restructuring their practice in

more fundamental ways characteristic of accommodation. It is also worth noting that teachers also responded to 27% of messages by rejecting them outright.

The degree to which teachers responded to state standards with superficial versus substantive implementation appears to depend, at least in part, on the level of capacity building that was available and that teachers took advantage of (Coburn, 2004; Dutro et al., 2002; McGill-Franzen et al., 2002; Spillane, 2000). For example, in an in-depth longitudinal study of one fifth-grade teacher, Spillane, Reiser, and Reimer (2002) used the contrast between reading and mathematics instruction to investigate the conditions that support change in practice. He provides evidence that this teacher made substantial changes in her practice in response to new English/Language Arts standards and argues that this was due, in part, to her active participation in the district's rich professional development offerings in reading instruction. Yet, this same teacher participated in only a limited way in professional development related to the new mathematics standards and made only superficial changes in her practice.

Nearly all the studies that investigate teachers' responses to state standards published during this period focus on implementation as learning. These studies emphasize the degree to which standards put forth visions of instruction that often require teachers to learn to teach in new and different ways. Accordingly, most of these studies drew on conceptual frameworks that attended to the dynamics of teacher learning and change. This focused attention on the nature of teachers' opportunities to learn about the instructional foci and approaches presented by standards and frameworks. These studies paid less attention to issues of power and authority, which are often central to studies of policy implementation outside of education. However, a few studies did investigate the way that school districts with authoritarian approaches to leadership created strong parameters within which teachers could learn and make change in their instructional approaches in response to standards (Dutro et al., 2002; McGill-Franzen et al., 2002). Overall, however, the contribution of this set of studies is to identify the learning demands that new reading policy placed on teachers and to investigate the dynamics by which teachers learn and make changes in their practice in response.

### *High Stakes Assessment*

The greatest number of studies during the review period focused on teachers' response to high stakes assessment. Taken together, these studies document teachers' negative attitudes toward standardized assessment. But they also show that in spite of these negative feelings, teachers did change their practice in response to high stakes testing. The nature of teachers' response depended upon the nature of the test.

Nearly all studies that investigate high stakes testing document teachers' widespread negative views. Teachers are especially negative about assessment when it is used for accountability purposes (Hoffman, Assaf, & Paris, 2001; Kauffman, Johnson, Kardos, Liu, & Peske, 2001; McDonnell, 2004; Wright & Choi, 2006). Several survey- and interview-based studies provide evidence that teachers question the degree to which standardized testing accurately measures student learning (Hoffman, Assaf, & Paris, 2001; McDonnell, 2004), especially in cases where the test was in English and students were predominantly English Language Learners (Hoffman et al., 2001; Wright & Choi, 2006). They are also skeptical that the emphasis on testing is beneficial for children (Hoffman et al., 2001).

Yet, at the same time that these studies document teachers' negative attitudes about testing, they also provide evidence that teachers do make changes in their practice in response to high stakes testing. Multiple studies—both survey-based and qualitative case studies—document increased or extensive test preparation activities (Diamond, 2007; Hoffman et al., 2001; McDonnell, 2004; Wright & Choi, 2006), narrowing of the curriculum to tested subjects and tested topics within tested subjects (Diamond, 2007; Hoffman et al., 2001; Sloan, 2006; Ogawa et al., 2004;

Wright & Choi, 2006), and redirection of instructional resources to “bubble kids”—those at the margins of proficient whose test scores could influence a school’s performance rating—at the expense of both high and low achieving students (Booher-Jennings, 2005; Diamond & Spillane, 2004).

Importantly, the ways in which teachers narrowed curriculum and the nature of test prep activities appears to be related to the nature of the assessment. For example, in his study of 47 teachers in 8 low-income schools in Chicago, Diamond (2007) provides evidence from 105 classroom observations that the city’s basic skills test tended to hold in place and reinforce didactic literacy instruction in spite of the presence of standards that emphasized higher order learning outcomes. This was especially true in schools that served large numbers of African American students.

In contrast, in her study of testing policy in North Carolina and Kentucky, McDonnell (2004) found that new state assessments focusing on open-ended prompts, collaborative work, student writing, and higher order outcomes resulted in teacher practice that moved towards these ends. McDonnell analyzed two weeks worth of daily logs of teacher practice and student work assignments for 46 teachers, coding them to assess the degree to which teacher practice and student assignments were consistent with the aims of the new state assessments. She found that teachers in the sample were adding reform practices consistent with state assessments to their instructional routines and that their student assignments reflected classroom activities associated with the assessment reform. However, in a finding that echoes the research on teachers’ response to standards reviewed above, McDonnell provides evidence that rather than fundamentally changing their pedagogical approach, teachers added new reform practice on top of existing traditional instruction. Teachers also were more likely to incorporate new activities (more writing assignments, more assignments that explored interdisciplinary connections) than to create instructional assignments that fostered conceptual understanding and critical thinking skills promoted by the testing reform.

Taken together, studies of teachers’ response to high stakes accountability provide evidence that teachers shift the nature of their instruction in the direction of the testing content. However, as McDonnell’s study illustrates, when the testing content required ambitious or unfamiliar approaches to instruction, teachers tended to change their practice in superficial ways and implement instructional approaches that addressed surface features of the policy rather than their deeper pedagogical implications.

### *Mandated Curriculum*

The final set of studies during the period of review focused on implementation of mandated curriculum. In most of the policy contexts investigated, mandated curricula were part of systems of instructional guidance that included regular interim assessments, pacing schedule, and intensive monitoring with a focus on fidelity of implementation. There are two distinct types of studies investigating this policy strategy: state-level Reading First evaluations (only a subset of which have substantive information on implementation)<sup>11</sup> and small-scale, close-in studies of teachers’ responses to mandated curriculum in Reading First and other policy contexts. These studies present a mixed picture about the degree to which teachers used the curriculum as intended and the changes in practice that resulted.

On the one hand, the large-scale evaluations of Reading First that attend to issues of classroom implementation report widespread presence of elements of Reading First, including the core curriculum (DeStefano, Hammer, & Fielder, 2006; Haager, Dhar, Moulton, & McMillian 2006; Zigmond & Bean, 2006). For example, in their evaluation report of the fourth year of Reading First in California, Haager and her colleagues (2006) draw on surveys of teachers, principals, and reading coaches to provide evidence that 98% of Reading First schools across

the state had “adequate” implementation of the provisions of Reading First and 4% had “better than adequate” implementation. Schools were considered to have “adequate” implementation if teachers, reading coaches, and principals rated their school as performing adequate on more than 50% of the items on an implementation index that included measures of participation in professional development, school level variables (for example, shared vision for reform), and self-report and coaches’ and principals’ report of classroom practice.

In contrast, smaller scale observational studies of Reading First and other policy initiatives that stress fidelity to mandated curriculum report that although most teachers were using mandated curricula, there was great variability and limited fidelity in their approach (Achinstein & Ogawa, 2006; Kersten, 2006; Kersten & Pardo, 2007; Sloan, 2006). For example, in his 3-year ethnographic study of three experienced teachers in one urban district in Texas, Sloan (2006) illustrates three entirely different responses to the district’s policy of mandated curriculum with explicit scope and sequence, quarterly benchmark assessments, and expanded battery of reading assessments. One teacher entirely ignored the curriculum, continuing with his largely project-based approach to teaching. A second teacher complained bitterly about the new curriculum and pacing schedule, but eventually integrated many key aspects of the curricula into her instruction over time. As her students’ performance improved, so did her attitude toward the instructional approaches embodied in the curriculum. A third teacher initially replaced many aspects of her existing reading program with the new curriculum and instructional approaches consistent with the assessments. However, after students’ perennially high test scores went down, she negotiated with the school leadership to allow her the flexibility to return to prior instructional approaches, which included little attention to the adopted curriculum.<sup>12</sup>

These small-scale, observational studies all suggest that when teachers do implement the curriculum, rather than implementing with fidelity, they tend to combine elements of the new curriculum with their prior instructional practice to create hybrid approaches (Achinstein & Ogawa, 2006; Kersten, 2006; Kersten & Pardo, 2007; Sloan, 2006), a finding that is consistent with studies of curriculum implementation that pre-date Reading First (Datnow, Borman, & Stringfield, 2000; Datnow & Castellano, 2000; Hoffman et al., 1998; Remillard, 2000, 2004) and studies of teachers’ response to state standards reviewed earlier. Indeed, the only observational study during the current period that provides evidence that teachers actually used highly specified curricula with a high degree of fidelity is a study of new teachers. In their longitudinal study of new teachers’ response to the curricula in their districts, Valencia, Place, Martin, and Grossman (2006) investigated 4 teachers, 2 of whom taught in districts with highly specified curriculum linked to pacing schedules and interim assessments. They document how these two teachers taught reading in ways largely guided by the mandated texts in what they characterize as a “procedural” rather than “conceptual” approach to instruction. However, the two teachers were the only ones across the four small-scale studies of mandated curricula that appear to follow the curriculum with a high level of fidelity.

These small-scale studies also document instances where teachers actively resist implementation (Achinstein & Ogawa, 2006; Kersten, 2006; Kersten & Pardo, 2007; Sloan, year).<sup>13</sup> For example, in their study of the role of school and district instructional policy on new teachers’ socialization, Achinstein and Ogawa (2006) found that 2 of the 9 novice teachers refused to use their school or district’s mandated curriculum as intended. In an article that profiles these teachers, the researchers argue that teachers did so because the curricula conflicted with their beliefs about high quality reading instruction and their view of themselves as professionals. Both teachers ultimately left the schools that required fidelity to the curriculum.

It is difficult to reconcile the different portraits of implementation put forth by these two different sets of studies. Large-scale evaluation studies report strong implementation. Small-scale observational studies document high degrees of variability, the widespread presence of hybrid practices, and some degree of resistance. On the one hand, the large-scale evaluation

studies have the benefit of drawing data from large numbers of teachers. They have large enough samples to run statistical tests that help them make generalizations and rule out spurious relationships. But, because of the scale of their studies, they are not able to employ very fine-grained measures of classroom practice. For example, Haager et al.'s (2006) study uses a wide range of items for their measure of Reading First implementation at the school level and, in fact, the items that are specific to classroom instruction play only a very small role in their implementation index (they are weighted at only 20% of the overall measure). They also rely on teachers' self-report and principals' and coaches' report of teacher practice rather than direct observation (see Rowan, Correnti, & Miller, 2002 and Hill, 2005 about the limits of surveys in accurately capturing teachers' instructional practice). Other large-scale evaluations of Reading First have measures of classroom implementation that go beyond surveys. However, these studies tend to rely on a very small number of observations—typically 2 a year per teacher—to draw their conclusions. As Rowan and colleagues (2002) suggest, this number of observations is problematic in light of the high level of variability in a given teacher's instruction across time.<sup>14</sup>

At the same time, the in-depth observational approaches favored by case studies provide a more fine-grained rendering of the nature of teachers' reading instruction, especially when the data come from extensive observations and/or are collected over several years. Thus, these studies may be surfacing aspects of instruction that are hidden from view with more general measures of instruction and reveal variation between teachers that is washed out by aggregation to the school or state level, as is the practice in the large-scale evaluations. However, most of the observational studies reviewed in this section provide limited information about their sampling strategy (Kersten, 2006; Kersten & Pardo, 2007) or purposively sampled outlier teachers (for example, Achinstein & Ogawa, 2006). Consequently, it is difficult to ascertain the degree to which the teachers profiled are representative of larger populations of teachers. Thus, while these studies provide evidence that there is variability in teacher practice, the presence of hybrid practices, and teacher resistance to mandated curriculum, we do not know how widespread these phenomena are.

Finally, in highlighting teacher resistance, the small-scale studies of teachers' response to mandated curricula highlight issues of power in the implementation process. Policy makers at multiple levels of the system are using various mechanisms of control—including mandates and monitoring—to ensure that teachers make changes in their classroom practice (in this case, adopt curricula). These articles highlight the fact that teachers do not always respond to this pressure; instead they reject some aspects of policy. However, at the same time that they shine a light on issues of power, studies of mandated curricula tend to neglect issues of learning that were highlighted in studies of standards and, to a lesser extent, standardized testing. That is, they pay little attention to what these policies require teachers to learn and the degree to which teachers have the knowledge and supports to do so. Indeed, most studies of mandated curricula assume that the teachers in the study fully understand the approaches to instruction promoted by the policy and could implement them with ease should they choose to do so.

### Summary

Taken together, studies across all three kinds of reading policy highlight several key findings. First, instructional policy does have the potential to influence teachers' classroom practice. However, when the policy promotes instructional approaches that are ambitious or unfamiliar, teachers are more likely to implement them in superficial ways rather than making fundamental changes in their instructional approach, especially when teachers have limited opportunities to learn about new approaches. Second, because teachers draw on their pre-existing practices to create hybrid approaches, there is likely to be great variability in teachers' implementation of reading policy, even within a single school. Third, studies of high stakes testing suggest



that teachers are making changes in practice in response to policy pressures even when the approaches are not in line with their beliefs about high-quality instruction. However, there is also evidence from studies of mandated curricula that under certain conditions, teachers resist policy pressures, although it is hard to discern how widespread this resistance is and when and under what conditions teachers resist policies that promote approaches to instruction that they themselves do not support.

The existing research has a series of methodological limitations. Some of the studies—both large scale and small scale—do not include enough observations of teachers' classrooms to ensure that they are adequately capturing the complexity of teachers' instructional practices or do not discuss changes in classroom instruction with enough precision to get a sense of the predominant or overarching pattern for a given teacher. Additionally, because most studies focus on a single policy at a time and because studies of different kinds of policy focus on different aspects of the policy implementation process, it is very difficult to ascertain the degree to which teachers respond to different policy strategies in different ways. For example, while it perhaps makes intuitive sense that there is greater incidence of resistance to Reading First and other policies centered on mandated curricula because of the aggressive forms of policy intervention they employ (i.e. mandates and extensive monitoring), it is not clear whether the incidence of resistance is in fact higher than in earlier periods. Coburn's study of teachers' response to a range of policy initiatives including state standards reported that teachers rejected 27% of all policy messages (2004). It is not clear if the incidence of resistance to mandated curriculum is higher, lower, or comparable to that figure. Similarly, it is not clear when teachers' limited response to state standards is an issue of learning (as is frequently the inference drawn in those studies) or an instance of resistance (an explanation that is largely absent in studies of standards).

### ***Factors that Influence Teachers' Response to Reading Policy***

In addition to documenting how teachers responded to changes in reading policy, some studies published during this time period also investigated the factors that influence when and how teachers respond in particular ways. In so doing, these studies begin to provide insight into the nature of policy implementation processes that produce policy outcomes. In this section, we draw on work by Spillane, Reiser, and Reimer (2002) to organize these findings. In his review of the intersection between policy implementation and cognition, Spillane argued that teachers' response to instructional policy is the result of the interaction between the individual (knowledge, beliefs, and attitudes), the social and organizational situation, and the policy signal (policy design). Taken together, these three categories provide a useful organizational framework for the explanatory factors highlighted in the studies of teacher responses to reading policy. We will address each set of factors in turn.<sup>15</sup>

#### ***Individual Factors***

In the last decade, studies of policy implementation have documented the way that individual teachers' prior beliefs, knowledge, and practices influence how they come to understand and enact instructional policy in their classrooms (Guthrie, 1990; Smith, 2000; Spillane, 1999; Spillane & Jennings, 1997; Coburn, 2001b & 2005). As Spillane and his colleagues explain: "Individuals must use their prior knowledge and experience to notice, make sense of, interpret, and react to incoming stimuli—all the while actively constructing meaning from their interactions with the environment, of which policy is part" (2002, pp. 393-394).

Indeed, several studies during this period provide evidence that teachers' *pre-existing knowledge and beliefs about reading instruction* play a major role in how they respond to new reading policies. These studies demonstrate that teachers' pre-existing beliefs and practices influence

what they notice about new policy, as they are more likely to focus on aspects of the policy that are familiar and not even notice those that challenge their beliefs and practices (Coburn, 2001b; Spillane et al., 2002). Pre-existing beliefs and practices also shape how teachers come to understand the meaning and implications of reading policy such that teachers with different pre-existing beliefs can understand the same policy in different ways (Coburn, 2001b, 2004; Spillane et al., 2002). Finally, prior beliefs and practice can influence change in classroom practice (Achinstein & Ogawa, 2006; Coburn, 2001b, 2004; Diamond, 2007; Spillane, 2000; Spillane et al., 2002). For example, in her study of three elementary teachers' response to changing reading policy in California from 1983 to 1999, Coburn (2004) provided evidence that teachers' responses to policy was shaped by the degree of congruence between new policy approaches and their pre-existing beliefs and practices. She shows that the greater the congruence of the policy message, the more likely it was that teachers incorporated the approach into their classroom practice in some manner. Thus, teachers responded to 90% of policy messages at high degree of congruence by creating parallel structures (e.g., a teacher adds a block of time for direct instruction of phonics skills and continues to teach phonics in a contextualized manner during small group instruction), assimilating them into their pre-existing practice (e.g., a teacher adopts learning centers but does not permit students to work collaboratively), or restructuring (i.e., accommodating) their practice in fundamental ways (e.g., a teacher learns about comprehension strategy instruction and fundamentally alters how she presents and discuss texts in the classroom). In contrast, the rate of incorporation dropped to 82% at a medium level of congruence and to 38% when congruence was low. However, although teachers were more likely to incorporate new approaches carried by policy with a high degree of congruence, they were also more likely to assimilate new approaches into their existing practice rather than reconstruct their practice in a substantive ways. In fact, teachers were most likely to make substantive changes in their practice in response to policy messages with a medium level of congruence. Just over 16% of the times that teachers incorporated messages at a medium level of congruence and nearly 9% at a low level of congruence, they did so in ways that led to accommodation. Messages with a high level of congruence led to accommodation in only 3 percent of policy messages.

The foregoing studies emphasize how the content of teachers' knowledge and practices shapes their response to policy. But the depth of teachers' knowledge is also important (Sloan, 2006; Valencia & Wixson, 2000). For example, Sloan's study of three experienced teachers with varied level of content knowledge in reading provides evidence that although mandated curriculum constrained the practice of the teacher with strong content knowledge, making it difficult for her to meet the diverse needs of her students, it improved the quality of instruction for the teachers with weak content knowledge (2006).

In addition to teachers' knowledge, beliefs, and practices, several studies provide evidence that *issues of identity* are also important for how teachers' respond to reading policy (Acheinstein & Ogawa, 2006; Kersten, 2006; Kersten & Pardo, 2007; Sloan, 2006; Spillane, 2000). Two aspects of identity are influential: a teacher's identity as a learner and his/her identity as a teacher. In his in-depth, longitudinal study of a single fifth-grade teacher, Spillane (2000) provided evidence that this teacher's identity as a learner influenced how she engaged in professional learning opportunities associated with new content standards and how that, in turn, influenced the changes she made in classroom practice. Recall from our earlier discussion that this teacher had opportunities to change both her literacy and mathematics practice. She engaged in multiple professional opportunities in literacy, but participated in hardly any related to mathematics across the 4 years of the study, even though there were plenty of opportunities available in the district. Spillane argues that the difference in participation was related to the teacher's identity as a learner, which varied substantially between literacy and mathematics. The teacher prioritized literacy in her teaching, saw herself as strong in this regard, and consequently actively sought out learning opportunities in literacy (including, but not limited to those provided by

the district). In contrast, she felt insecure about her abilities in mathematics, did not like teaching mathematics, did not see mathematics as central to her teaching mission, and consequently did not engage as actively with available learning opportunities in mathematics. As a result, this teacher made substantial changes in her practice towards the ambitious goals set forth by the literacy standards, she made only superficial changes in her mathematics teaching.

Other scholars focus on how teachers' identity as professionals—their sense of their appropriate role and what it means to be a teacher—influenced their response to reading policy. This body of work, most of which focus on teachers' responses to mandated curricula, suggest that teachers resist following well-specified curricula when this requirement conflicts with their view of themselves as a professional (Achinstein & Ogawa, 2006; Kersten, 2006; Kersten & Pardo, 2007) or when they do not identify strongly with the world of schooling (Sloan, 2006). For example, the two novice teachers who actively resisted the mandated curriculum profiled by Achinstein and Ogawa (2006) had visions of teaching as involving autonomy, creativity, and individuality. Following a highly specified curriculum conflicted with their view of themselves as teachers. Both teachers rejected using the curriculum as intended and ultimately left the schools that required them to do so.

### Situation

Individual characteristics shape teachers' responses to policy but so do the complex social and organizational contexts in which they work. Several facets of these contexts matter for how teachers implement reading policy in their classroom: social interaction with colleagues, school leadership, and features of the district.

Several studies suggest that teachers' *social interaction with their colleagues* influences their decisions about how to implement new approaches to reading instruction. First, patterns of social interaction influence how teachers learn about reading policy in the first place. Reading policy interpenetrates schools to different degrees. There is great variability in the degree to which teachers have access to information about new policy initiatives, especially if it is created at higher levels of the system (Coburn, 2005a; McDonnell, 2004). As suggested by research on social networks (Granovetter, 1973, 1982; Burt, 1992), teachers' connections with each other and those outside the school provide a powerful mechanism for learning about new ideas and approaches promoted by new reading policy, shaping teachers' access to some policy messages and not others (Coburn, 2001b, 2005a; Coburn & Stein, 2006). Second, teachers' interactions with their colleagues influence how they come to understand the meaning and implications of a new policy. Teachers make decisions about how to respond to new policy initiatives in conversation with their colleagues. How they come to understand the nature of instructional change required by a given policy is shaped by who they are interacting with (Coburn, 2001b; Booher-Jennings, 2005). For example, Booher-Jennings (2005) draws on extensive interviews and observations in a single school to provide evidence that teachers responded to high stakes tests by rationing instructional resources—focusing on children at the margins of proficiency (the “bubble kids”) rather than high-performing or low-performing students—in part because they felt pressure from colleagues to improve the schools' performance rating. Thus, teachers influenced each other to see rationing as the most appropriate response to the district's high stakes accountability policy.

Beyond social interaction, studies during this period highlight the role of *school leadership* in classroom implementation. Principals influence classroom implementation by emphasizing some aspects of reading policy and not others, shaping teachers' access to some aspects of reading policy and not others (Coburn, 2001b; 2005b; Diamond, 2007; Diamond & Spillane, 2004; Spillane et al., 2002). At a deeper level, school leaders also influence how teachers come to understand the meaning and implication of policy (Anagnostopoulos & Rutledge, 2007;

Coburn 2001, 2005b, 2006). Work by Diamond and Spillane (2004) suggested that school leaders' choices about what to emphasize, in turn, are influenced by the performance level of the school. Drawing on interview and observational data from a 4-year study of four low-income schools in Chicago—two high performing and two under threat to probation—Diamond and Spillane provide evidence that many of the practices noted by researchers in earlier studies of high stakes accountability—curriculum narrowing, focus on “bubble kids,” extensive test preparation activities, symbolic responses—were more likely to happen in low-performing low-income schools than high-performing low-income schools. School leaders in high-performing schools responded to high-stakes assessment by focusing attention on long-term improvement in all curricular areas (rather than just tested subjects), encouraging teachers to participate in test preparation activities just prior to the tests rather than all year long, and leading teachers to be more reflective and purposeful about instructional change efforts than schools under threat of probation. These studies go beyond the platitude that school leaders are important for instructional reform to begin to unpack the relationship between specific actions by school leaders and teacher learning and instructional change.

Finally, scholars have increasingly identified *features of the school district* as important for teachers' implementation of reading policy. School districts influence implementation by the level and quality of capacity-building activities (coaching and professional development) that they provide to teachers to support instructional change (Dutro et al., 2002; McGill-Franzen et al., 2002; Stein & D'Amico, 2002). Several authors also provide evidence that the degree to which teachers respond to state standards in superficial versus more substantive ways was influenced by the nature of district standards themselves (McGill-Franzen et al., 2002; Ogawa et al., 2004). For example, Ogawa and his colleagues draw on longitudinal data from their study of one California district to investigate the relationship between district standard setting and teachers' classroom practice. They document how the district created three levels of standards, all of which were actually lower than the state standards because district administrators felt that the state standards represented learning outcomes that were too challenging for the children in the districts. Ogawa and his colleagues provide evidence from teacher surveys that teachers in elementary grades tended to pitch their instruction to the lowest level of standards in the district, which focused on basic skills instruction. Interviews with teachers suggest that this focus on basic skills represented a change in practice in response to standards and the accompanying district test rather than a continuation of their existing instructional practice. This suggests that the district's standards and assessment practices mediated state standards, influencing teachers to focus on lower level outcomes rather than higher order outcomes promoted by the state standards.

A study by Booher-Jennings (2005) suggested another way that school districts influence implementation, in this case teachers' response to high-stakes accountability. She argued that teachers consented to engage in rationing of educational resources (e.g., devoting resources disproportionately to the “bubble kids”) in part because of their fears of being perceived as a “bad teacher” by their colleagues. The district contributed to this fear by defining good teaching by test score alone and publicizing individual teachers' test scores. Teachers' colleagues enforced this definition of good teaching, creating normative pressure that guided teachers' instructional choices. Taken together, these studies suggest that school districts influence teachers' implementation directly by creating instructional expectations in the form of standards and providing opportunities for teachers to learn new approaches, but also indirectly by creating normative environments and definitions of high quality teaching.

### *Policy Signal*

The final set of factors we will consider are those that relate to the nature of the policy message itself. While implementation researchers outside of education have placed a priority on

understanding the role that the design of policy plays in the nature of implementation (Matland, 1995; Mazmanian & Sabatier, 1983; Pressman & Wildavsky, 1984), the nature of the policy itself has received less attention in educational research. Indeed, few of the articles we reviewed on reading policy paid explicit attention to this dimension at all, in spite of the fact that the different policy movements during the last decade used substantively different policy strategies for influencing teacher behavior, resulting in very different policy signals. Nonetheless, there are a few patterns that do emerge from a careful reading of the recent research.

First, several scholars provide evidence that the *degree of ambiguity* of the policy influences how teachers implement reading policy. They provide evidence that when reading policies are ambiguous, teachers struggle to implement in the absence of adequate guidance (McDonnell, 2004; Kauffman, Johnson, Kardos, Liu, & Peske, 2002). For example, in her study of teachers' response to high stakes testing in California, Kentucky, and North Carolina, McDonnell provides evidence that testing policy, while providing a sense of urgency for instructional change, provided little guidance for how teachers should go about improving instruction to produce stronger student outcomes. Drawing on surveys and an analysis of classroom tasks (described earlier), McDonnell provides evidence that teachers had only a "diffuse, shallow understanding of assessment policy goals" (p. 156). Similarly, in their interview-based study of 50 first and second year teachers in Massachusetts, Kauffman and his colleagues (2002) found that in spite of the presence of state standards and assessment, beginning teachers reported that they had very little guidance about what and how to teach. All but a few teachers in their sample reported that they had either no curriculum at all (one-fifth of teachers) or only a list of topics and skills they needed to cover (over one-half) to guide their instruction. In these instances, new teachers struggled to build a curriculum on their own by piecing together materials from multiple sources, often without guidance from their school or district.

Research and theory outside of education suggests that increased ambiguity leads to greater variability in implementation. The basic argument is that the more ambiguous a policy, the greater opportunity for implementers to interpret and enact the policy in a wide range of ways leading to variability in implementation (Mazmanian & Sabatier, 1983). However, our earlier review of teachers' response to reading policies with different degrees of ambiguity (with standards being the most ambiguous and mandated curriculum accompanied by monitoring being the least) found evidence of variability in implementation across all policy strategies. In the absence of studies of that investigate the impact of different policy designs comparatively, and in the absence of common measures of implementation across studies of different policy initiatives, it is difficult to assess whether variation in policy ambiguity leads to greater or lesser variability in teacher implementation.

At the same time, at least one study provides evidence of the downside of too much specification. In their longitudinal study of four new teachers' engagement with reading curricula, Valencia and her colleagues (2006) draw on extensive interviews and repeated observations to provide evidence that the two teachers who worked in schools that had little specification in their curriculum struggled initially but eventually developed much greater understanding of reading instruction, greater flexibility, and more sophisticated instruction over the first three years of their teaching career. In contrast, the two teachers who started their career with highly-specified curricula had a more procedural and less conceptual orientation toward reading instruction and made few changes in their approach to reading instruction during their first three years of teaching. Thus, high specification ensured more support, but also offered fewer opportunities for teacher growth.

Second, several studies provide evidence that the *degree of alignment* among the multiple reading initiatives that teachers experience simultaneously influences implementation. These studies acknowledge the fact that teachers rarely experience a single policy initiative in isolation (Coburn, 2005a; Diamond, 2007; Kersten & Pardo, 2007). In the presence of multiple and



at times conflicting reading policies, teachers tend to pick and choose which policy messages to be responsive to and which to ignore. In making these choices, they were more likely to be responsive to policy messages that are more consistent with their beliefs and prior practice. For example, Kersten (2006) painstakingly documented how one experienced third-grade teacher constructed her response to Reading First. Faced with a district curriculum guide that was not aligned with Reading First mandates, the teacher opted to be responsive to district policy when it was consistent with her prior practice. In this way, the presence of multiple and conflicting policies legitimized this teacher's lack of fidelity to some of the tenets of Reading First.

Third, several scholars focus on the degree to which reading initiatives provide *capacity building elements* as part of their policy designs (Coburn, 2005a; Dutro et al., 2002; McDonnell, 2004; Stein & D'Amico, 2002). These studies suggest that opportunities for teachers to learn new instructional approaches are crucial if teachers are to make substantive changes in their classroom practice. For example, in their study of New York City District 2's comprehensive literacy policy, Stein and D'Amico (2002) provide evidence that teachers with higher levels of participation in the district professional development offerings had deeper enactment of the balanced literacy program. More specifically, Stein and D'Amico made 27 observations of 12 teachers, analyzing their instruction along two dimensions: alignment with the balanced literacy program and quality of their instruction. They show that those teachers who had participated in the district's extensive and high quality professional development activities for multiple years tended to have instruction that was both aligned with the policy and of high quality. In contrast, those with more limited experience with professional development tended to have high alignment, but low quality implementation. That is, while the teachers included the various activities associated with the balanced literacy approach in their classroom (high alignment), they implemented them in ways that did not reflect the underlying pedagogical principals that knit the activities together (low quality). The implication is that participation in high quality professional development over an extended period of time enabled teachers to move from more superficial to more substantive enactment of policy.<sup>16</sup>

However, Spillane's study comparing one teachers' response to mathematics versus literacy standards (2000) complicates findings about capacity building. Recall that the teacher in his study worked in a district that had extensive professional development opportunities in both reading and mathematics. Because of differences in this teachers' identity as a learner in literacy versus mathematics, she actively participated in professional development in literacy, but hardly participated at all in professional development in mathematics. This difference was reflected in deep changes in this teachers' literacy instruction compared to only superficial changes in her mathematics instruction. This difference suggests that capacity building efforts may be necessary but not sufficient to support substantive implementation of standards. Rather, there may be an important interaction between elements of the policy, like the nature and quality of mechanisms for capacity building, and features of the individual teacher, such as inclination to reach out and engage with the opportunities that are available.

During the period under review, attention to issues of capacity building as part of policy design was largely confined to investigations of state standards and, to a lesser extent, high stakes assessment. In spite of the fact that Reading First and many other similar policies involve extensive capacity building activities, few studies of mandated curriculum attended to this issue. A few state evaluations of Reading First report data on teachers' and principals' satisfaction with coaching or professional development (Haager et al., 2006; Zigmond & Bean, 2006), but none look at the relationship between the levels and quality of capacity building activities and changes in teachers' classroom practice.

Finally, Coburn (2004, 2005a) provided evidence that the *degree of voluntariness* of policy influences how teachers respond to reading policy. Coburn used a longitudinal design, which provided the opportunity to contrast teachers' responses to multiple reading policies, each of

which employed quite different policy strategies. Coburn distinguished between policy that uses normative pressure, putting forth visions of high quality instruction and making arguments for why teachers *should* make changes in practice (for example, standards) versus those that employ regulative pressures, which require teachers to do particular things and enforce this vision of instruction using rules, monitoring, and sanctioning (for example, mandated curricula). Coburn documented how, during the time period of her study, a vast majority of policy messages that teachers encountered were normative than regulative. Although teachers were less likely to reject policy messages accompanied by regulative pressure, they were also less likely to incorporate messages accompanied by regulative pressure into their classroom in substantive ways than they were to incorporate those messages offered with normative pressure. Thus, teachers responded to 33.3% of regulative messages symbolically (a kind of mock compliance), 20.8% of regulative messages by assimilating them into their pre-existing instructional approach, but never responded to regulative messages by reconstructing their practice in fundamental ways. In contrast, teachers responded to only 4% of normative messages symbolically, 51.6% by assimilating messages into pre-existing practices, and 10.6% by reconstructing their practice in fundamental ways. It is important to note that data collection for Coburn's study was completed in 1999 before the advent of high stakes testing and mandated curriculum in California. It will be important to see how these findings hold with policy initiatives that make much more extensive use of regulatory pressures than those investigated by Coburn.

### Summary

Taken together, these studies highlight the complex web of factors that influence how and why teachers respond to reading policy in particular ways. They paint a portrait of individual teachers making decisions about their practice in ways that are guided by their history, identity, existing knowledge and practice, but also influenced by the nature and quality of their interaction with their colleagues. The broader school and district context play a role as well, shaping teachers' access to some policy ideas and not others, creating opportunities for teachers to learn, and creating normative expectations for good teaching that shape how teachers see themselves and their colleagues. Finally, these studies provide evidence that the nature of the policy message itself matters, its ambiguity, alignment, instruments for capacity building, and its voluntariness. However, there is still much to learn. To date, scholars of reading policy have had a tendency to focus greater attention on how teachers respond to new policy initiatives than to dig in and uncover the factors that shape these responses. Those that focus on why teachers respond in particular ways tend to emphasize individual level factors, such as beliefs and identity. As we move from individual to contextual and to features of the policy itself, there are fewer and fewer studies that have taken these factors into account.

Similarly, initial studies have identified a plethora of factors that influence implementation. But few studies have systematically investigated how these factors interact to influence classroom practice. For example, studies suggest that degree of specification and ambiguity influences the nature of implementation. We also know that teachers' depth of content knowledge may matter for how they respond to various kinds of policy. However, there are no studies that investigate how degree of specification and ambiguity interact with teachers' pre-existing content knowledge. Do teachers with different levels of content knowledge respond in different ways to highly specified policies? To more ambiguous policies?

Systematic studies of this sort are hampered because so few researchers compare implementation processes across different kinds of policies or similar policy in different contexts. Because most studies of reading policy are cross-sectional, they study teachers' response to a single kind of policy (in this case, standards, high stakes assessment, or mandated curriculum) rather than compare teachers' response across policies with different features (Coburn, 2004, 2005a are

exceptions). Similarly, most studies of reading policy investigate teachers in a single district or a single state (Dutro et al., 2002 and McDonnell, 2004 are exceptions). This makes it difficult to make systematic comparisons about how policy context matters for implementation. It also loses the opportunity to investigate the consequences for classroom practice when states or districts implement a single policy—like Reading First—in different ways. In the absence of designs that compare different kinds of policies or similar policies in different contexts, it is very difficult to untangle the role of situation and policy signal in classroom implementation. This, in turn, makes it challenging to investigate how these kinds of factors interact with teacher- or school-level factors to produce patterns in classroom implementation.

As the field matures, it will be important to push our understanding of the factors that influence teachers' implementation further. At a minimum, we need more studies that seek to go beyond documenting how teachers responded to uncover the factors that influence why teachers respond in particular ways. But we also need the development of more precise measures of practice to facilitate comparison of classroom practice across multiple studies. And we need more studies that are designed to facilitate systematic comparisons—between kinds of policy designs, kinds of settings, and teachers with varied beliefs, knowledge, and practices—such that we are able to understand the interactions between the multiple factors that influence classroom implementation of reading policy.

## POLICY OUTCOMES: FOCUS ON READING FIRST

In this section we take advantage of a singular opportunity provided by the array of national and state evaluations of federal programs—the Reading First component of No Child Left Behind—over the last decade. Never in the history of reading instruction have we had so many resources focused on early reading achievement, and, as a result, so much evaluation and research evidence gathered in such a concentrated period of time about systematic efforts to reform reading instruction in our schools. The various studies vary in purpose, method, findings, and conclusions, leading observers to conclude that Reading First has been the “Rorschach test” of literacy policy in the United States, or perhaps something more akin to the proverbial elephant described in such contradictory terms by the various blind observers. Even so, there is much to be learned from these evaluations, some of it substantive, but some of it methodological and even ideological.

For this analysis, we took a strategic approach in sampling the overwhelming array of available documents. First, we took careful note of the two large national quasi-experiments on Reading First: the *Reading First Impact Study Final Report* (Gamse et al., 2008) and the *National Evaluation of Early Reading First* (Jackson et al., 2007). Second, we relied on the documents assembled on the website of the Center for Educational Policy ([www.cep-dc.org](http://www.cep-dc.org)), including their own evaluations of Reading First and their summaries of other work, as well as their links to evaluations conducted by the federal government and other agencies. Of remarkable assistance was the 2007 aptly titled CEP report, *Reading First: Locally Appreciated, Nationally Troubled* (Scott, 2007), which summarized much of the state evaluation data up to that point in time. Third, we made the strategic decision that we could not examine all of the state reports (archived at <http://www2.ed.gov/programs/readingfirst/evaluationreports/>), so we decided to sample state evaluations of Reading First on a principled basis, looking for both within and between state patterns. We selected 5 states with the largest Reading First Grants (California, Texas, Florida, Pennsylvania, and Illinois) on the grounds that they were the grants that affected the largest number of children and schools in the country. We brought in additional findings and perspectives from the reports from other states when they served to reinforce or extend findings from the 5 large states or provide an alternative perspective.

### **The National Reports**

The two national evaluation studies—the *RF Impact Study* (Gamse et al., 2008) and the *National Evaluation of ERF* (Jackson et al., 2007)—were explicit attempts on behalf of the Institute of Education Sciences (IES) to come as close as possible to meeting the gold standard of a randomized field trial in evaluating the effects of RF and ERF. True random assignment was not possible because of the manner in which funds were allocated. Because the highest rated proposals from eligible districts within each state received the funding, the assignment of treatment (RF or not) to the unit of analysis could not be random. However, in such circumstances, a common alternative to random assignment is the regression discontinuity design (Trochim, 1984). In such a design, under the assumption that the funded units differed from the unfunded units *only on* the variables that underlie the “scores” used to allocate the funding, the scores received on proposals are used to adjust outcome scores in the analysis. The closer the experimental and comparison groups are to the cut score, the better. Thus sampling from units just above and just below to cut score is a common approach to establishing both an experimental and a comparison group. This approach was used in both the *RF Impact Study* and the *Evaluation of ERF*. In what has proven to be a controversial move, the designers of the RF evaluation applied the design tool within districts (selecting schools just above and below the local cut points for school level funding), causing some scholars (e.g., Reading First Advisory Committee, 2008) to worry about between-group contamination of policy and practice initiatives.

The findings in *RF Impact Study* (Gamse et al., 2008) are straightforward. Differences favoring RF schools were found on a number of program implementation variables—total time spent on reading and practicing the “big five” RF components, explicit instruction (grades 1 and 2), high quality instruction (Grade 2 only), time spent on reading (hours per day), focus of professional development on the big five components, and the effective deployment of reading coaches. Implementation differences were not found on student engagement with print, access to differentiated instruction, or the use of diagnostic assessment. On student outcome measures, differences ( $ES$  of .17) were found favoring the RF schools on a measure of decoding skill in Grade 1 but not on comprehension at any grade. However, these findings are also controversial because they are inconsistent with the state evaluations of RF, most of which demonstrated robust effects favoring RF schools on a range of outcomes (cf. p. 49).

The *Evaluation of ERF* was much less ambitious in scope than the *RF Impact Study*, paralleling the scope of the funding differentiating the two programs. Like the larger study, the ERF study employed a regression discontinuity design with the school site as the unit of analysis, but the unit of selection, as would be expected with pre-school programs, was either a stand-alone pre-school or a “consortium” with several constituent pre-school sites. For both ERF and comparison treatments, classrooms were randomly selected within school sites for participation. The findings also parallel those for the *RF Impact Study*: strong effects on measures of program implementation and weak effects on student outcomes. In particular, ERF demonstrated positive program implementation effects on (a) teachers’ professional development opportunities, particularly related to language development and literacy; (b) teachers’ sensitivity to children’s needs ( $ES = .79$ ); (c) quality of teachers’ interactions with students; (d) classroom organization and quality of learning environment; (e) lesson planning; and (f) the relative emphasis during literacy lessons on a range of practices: oral language, book reading, phonological awareness, print and letter knowledge, writing, and screening devices. Of the three classes of outcome measures—print awareness, phonological awareness, and oral language, effects favoring ERF were found only on print awareness. The evaluators also examined the mediating effect of changes in teacher and classroom variables on student achievement, finding a reliable (and predictable) effect only for print and letter knowledge activities on print and letter knowledge achievement.

Looking across the two national studies, a consistent message seems to be that when rigorous quasi-experimental designs are used to evaluate the impact of the complex, multi-faceted programs, it is easier to demonstrate reliable effects on measures of teacher participation and practices than on student outcomes. In this regard, it is worth noting that Dee and Jacob (2009) found, in their broader analysis of the impact of NCLB on student achievement as reflected in a set of broader indicators (scores on fourth- and eighth-grade NAEP), that the evidence for NCLB's influence on mathematics performance is quite strong, while the evidence for impact on reading scores is essentially non-existent.

### ***State Evaluations of Reading First***

With some notable exceptions, namely the relatively flat and disappointing results from Texas, the state level evaluations show consistently positive programmatic and student outcomes. The consistent message across the state reports is that Reading First worked, at least as measured by the criteria used to evaluate its impact on school programs, professional development, teacher practices, and student achievement (Carlson, Branum-Martin, Durand, Barr, & Francis, 2008; Foorman, Petscher, Lefsky, & Toste, in press; Haager, Dhar, Moulton, & McMillian, 2008; MGT of America, 2008). By and large, teachers appreciated the various programs (all variations on the "big five" theme—phonemic awareness, phonics, fluency, vocabulary, and comprehension—from the NRP Report), liked and responded well to the staff development provided, and implemented the key components of the enabling legislation.

Even more important, students appeared to benefit from the programmatic changes that were implemented. The general trends, using measures that focused on comprehension reported across these five states, indicate that:

1. Students in Reading First schools outperformed the comparison group, where the comparison group was defined as either as a comparison group of schools (in California),<sup>17</sup> the large sample used to establish the norms for the commercial tests (the SAT in Florida or the ITBS or SAT in Texas) or the achievement level cut scores<sup>18</sup> for the statewide accountability measures in Florida (Grade 3) and Pennsylvania (Grade 3).
2. The Reading First advantage extended to the lowest achieving students. Many states noted substantial reductions in the percentage of students scoring below their "basic" cut scores (or the 25th percentile or 20th percentile for the states using commercial tests). An especially gratifying result is that Florida schools that did not make progress early on were given special assistance in the form of site visits and special coaching and, as a result, were able to accelerate their progress in subsequent years.
3. With a few notable exceptions, traditionally underachieving groups (low-income students, students with disabilities, or ethnic or linguistic minority students) made exceptional progress in comparison to students in the comparison populations. In fact, in some states (California, Florida, and Pennsylvania are the best examples), there was some evidence of actually closing the achievement gap, although it must be recognized that RF seemed to benefit students at all levels of prior achievement and demographic factors.
4. In general, time in the program (the more years students were part of Reading First, the greater the gains), implementation fidelity (the higher the level of implementation, the greater the gains across time), and mobility in both students and teachers (the greater the mobility, the lower the rate of progress) explained variance in school achievement after many other factors, including SES, had been controlled for.



***Reconciling the Differences Between State and National Evaluations***

So what are we to make of this impressive array of state level findings, all pointing toward the overall efficacy of Reading First, in light of the singularly unimpressive results of the national studies? There are several plausible explanations, all of which should be considered. First, the RF effect emerges in less rigorous designs, such as evaluations that compare RF to population norms or non-comparable comparison groups, but not in more rigorous designs, such as randomized experiments or regression discontinuity designs. It could be, for example, that there are unaccounted for and unintended confounds between RF and non-RF groups in the state evaluations. These confounds could be stem from many potential factors—attention, resources, or intellectual capital that creates a special advantage for an experimental group. If this argument prevails, then we should look beyond the programmatic elements of RF for answers to our questions of what works.

Second, it is logically possible that the RF effect in the evaluation studies is real, and it is the *Reading First Impact Study* and the *Evaluation of Early Reading First* that suffered contamination between treatments. This is exactly the position taken by the National Reading First Advisory Committee (2008) when they saw the draft of the Impact Study. Their concern was that the decision to situate both experimental and control schools within districts and then to allow districts to assign treatment to schools compromised the integrity of the design. They further point out that many districts, in trying to build district capacity for RF, may have unintentionally contaminated the culture(s) of the control schools with ideas and principles emerging from the RF programmatic and professional development efforts. Finally, the National Reading First Advisory Committee expressed a concern about external validity—that the decision to embed the study within districts that used some sort of “rank-ordering” procedure for assigning RF treatment to schools may have inadvertently limited the sample to a restricted subset of schools, thus rendering the sample unrepresentative. If this argument prevails, then we should be asking ourselves questions like: Now that we know that these programmatic elements matter, what else do we need to add to our curricular and professional development portfolios in order to take the next step in reform and innovation?

A third possibility is that the RF effect is real, but not for any of the reasons typically cited by those who support its policy status. This is a variation of the “reading drive” argument put forth long ago by Southgate (1966). The idea is that novelty matters. Once in a while, change is needed to stir things up, offer participants hope and promise, and motivate everyone in the setting to improve. There is a hint of this perspective in the thoughtful set of conclusions offered by the authors of the Pennsylvania report (Bean, Draper, Turner, & Zigmond, in press). Bean and her colleagues conclude that the focus RF provided across Pennsylvania—along with local variation in implementation strategies— was important for the outcomes they observed. Citing McLaughlin’s (1976) concept of mutual adaptation between reform and local school culture as an explanatory factor, they argue that strong outcomes in Pennsylvania were due in part to the fact that local districts differentiated their implementation strategy to meet their local needs and capacities, but kept this variation within the framework of RF dictates. In so doing, they were able to achieve implementation that led to strong student outcomes. If this argument prevails, then we should probably look to more open and process-focused approaches to reform—interventions that offer participants prerogative in shaping new initiatives (see Taylor, Pearson, Peterson, & Rodriguez, 2005, or Levin, 1998) within a common framework.

Ultimately, we do not have enough information to know which of these possible interpretations of the conflicting findings is the most accurate. This state of affairs highlights several limitations—methodological and conceptual—in existing research on the outcomes of reading policy in general and Reading First in particular. First, few of these studies use multiple measures of student achievement. As Paris (2005) has argued, it is preferable to have multiple

measures of student achievement that range from low-level micro-level enabling skills (e.g., phonemic awareness, phonics, or word identification) to more macro-level outcomes (e.g., comprehension and language development). Doing so allows the researcher to understand the range of phenomena that interventions do or do not influence; a single measure cannot achieve that goal. Second, few of these studies collect measures on pedagogical practice. Absent information about the nature and quality of instruction, it is very difficult to make sense out of achievement results. Without this information, we have little information about the “active ingredients” in these classrooms and interventions.

Third, while all the evaluation studies of Reading First used comparison groups, many of the comparisons were problematic. Comparing the growth of the experimental group with changes in the overall state population or the average performance of the norming group for a commercial assessment, as some state evaluations do, is a poor substitute for a randomly selected control group since we know next to nothing about (a) the natural and unintentional distribution of ingredients of the intervention in a large and diverse norming groups or (b) the demographic similarity of the experimental and norming group. Even more typical control groups can be problematic on two oppositional counts. First, for an intervention like Reading First, contamination is likely to be the norm not the exception. As the California report documented, Reading First was a pervasive reform, extending to fourth and fifth grade in most schools and even to schools that did not receive direct funding; as such there was a natural press for contamination and conceptual seepage in the state’s educational reform culture (Haager et al., 2008). Second, business as usual control groups are seldom treated comparably on important and potentially confounding factors such as material resources, professional development opportunities, and attention—all factors that might motivate teachers to teach in ways that would lead to higher achievement quite independent of the content or focus of the interventions. Finding ways to control for these almost inevitable differences between treatments poses a genuine dilemma for researchers. Developing comparison groups that represent viable alternatives to the intervention under study is one strategy, albeit an expensive one. When random assignment is not possible, as is often the case in school-based evaluations, some scholars advocate for regression discontinuity designs, such as employed in the *Reading First Impact Study* (Gamse et al., 2008) and the *National Evaluation of ERF* (Jackson et al., 2007), but even they introduce vulnerabilities as discussed above (Reading First Advisory Committee, 2009).

We know why most evaluation studies fail to meet these ideals: resources. Evaluation studies rarely have the resources they need to answer the questions they have asked. However, sometimes we need to sacrifice breadth of coverage (testing the entire sample) to achieve greater analytic depth (more measures of more variables, including contextual variables such as instruction). The reallocation of resources in this manner may make it more likely that evaluations can improve our understanding of the impact of ambitious reading policy on student outcomes.

## CONCLUSIONS

As policy makers become more and more interested in using policy to influence instruction, it becomes increasingly important to attend to the causes, processes, and outcomes of these efforts. Fortunately, research on reading policy is burgeoning as well. And, we are beginning to learn some crucial lessons about the relationship between reading policy, teachers’ classroom practice, and student outcomes. We know that research can influence the development of reading policy, but that influence is dependent upon the structure of the policy issue networks that bring ideas and approaches into the hands of policy makers. We also know that these issue networks can change substantially over time.

We know that various approaches to instructional policy does reach within the classroom door to influence classroom practice. But, if policy is ambitious or unfamiliar, teachers are likely to implement it in superficial or tangential ways in the absence of capacity building efforts. We know, too, that there is likely to be great variability in the ways that teachers implement policy, even within a given school. Teachers will make change in their practice, even when they do not support the instructional approach promoted by the policy. But we also know that teachers can and do actively resist policy, picking and choosing among the plethora of policy messages to implement those that most resemble their pre-existing beliefs and practices. Understanding which of these implementation outcomes happens under what conditions is the next frontier for research. And, indeed, existing research has identified a web of factors—some individual, some social, some contextual, and some features of the policy itself—that influences how teachers respond to policy in their classrooms.

However, even as the research on reading policy begins to develop, there are still notable gaps in our understandings. There is still very little research that investigates the dynamics of policy making. We know little about how issue networks form, how they change over time, and how reading research and researchers become key players in them. This understanding is critical for reading researchers because the more we understand the process by which research moves in and through the system and into policy, the more we can work to form a more productive relationship between research and policy making in the service of improving reading instruction.

Furthermore, in spite of an increase in studies of the process of policy implementation, there is still much to learn about how policy influences teachers' classroom practice. We are beginning to identify the factors that shape implementation of policy, but we know little about when and under what conditions a given factor is important or how they might interact to produce the changes in practice that we see. Existing research has tended to focus on individual factors (i.e., beliefs, knowledge, identity) and paid less attention to organizational context or features of the policy itself. In fact, there are almost no studies that compare different policy approaches to discern how different policy tools (mandates versus capacity building versus accountability mechanisms) influence classroom practice differently. This, in spite of the fact that there is some variability in the same policies across different states (e.g., NCLB and Reading First) and even different districts that would facilitate this sort of comparison. Understanding the implementation process is crucial if we are to open the black box between a policy and its outcomes. Doing so promises to provide insight into strategic levers to support and sustain instructional improvement.

Finally, in spite of increased funding for evaluations of Reading First, there continues to be a paucity of funding for systematic studies of policy outcomes. For many researchers and policy makers, student learning is the bottom line. Yet, there are many significant reading policies with virtually no studies of student outcomes. The Reading Excellence Act, a major piece of federal legislation on reading instruction, received remarkably little research attention. Similarly, there are many state-level policies (e.g., new credentialing requirements for reading teachers in several states) for which we have no systematic analysis of outcomes.

Furthermore, studies of policy implementation and policy outcomes alike are hampered by methodological limitations related to measuring instructional practice. Many studies—especially the large-scale evaluations—have made limited use of classroom observation and other low-inference measures to investigate how teachers are actually responding instructionally. Those studies that do have rich measures of classroom instruction often use unique or idiosyncratic approaches, making comparison across studies difficult. The small-scale studies, which typically have more extensive observational components, often provide little information about sampling choices, so it is difficult to situate the classrooms and schools in the study in the larger sample. All of this makes it difficult to identify just how policy is impacting classroom practice and what features of instruction are associated with increased student learning. If we are truly to

understand how policy impacts student outcomes, we need studies that help us understand what is going on instructionally to produce those outcomes. Absent that attention, policy makers have little information to guide the development of policy tools to promote and support efforts to improve reading instruction and student learning.

## NOTES

1. While this was the first attention to “scientifically-based” approaches in federal legislation, references to scientifically-based research had appeared in state legislation in California as early as 1996 (Coburn, 2001a; Pearson, 2004) and in Texas by 1997 (Texas State Education Agency, 1997).
2. Student ranking in high school also included scores on a writing, social studies, and science tests.
3. The Texas model later became the blueprint for No Child Left Behind, as key Texas officials took positions in the Department of Education during the Bush administration.
4. It is perhaps interesting to note that the mandate for states to adopt scientifically-based research proved to be fraught with difficulties, as officials at the federal level were found to have pressured states to choose particular materials, some of which were authored by Department of Education officials or advisor. On March 23, 2007, the Government Accountability Office (2007) issued a report corroborating the findings from six reports issued by the U.S. Department of Education’s inspector general. In particular, it found that “federal officials failed to safeguard against potential conflicts of interest in administering the program; and they directed some states’ and districts’ choices of reading texts and assessments, despite legal prohibitions” (Manzo, 2007). In particular, Reading First officials were accused of privileging programs or tests in which they or close associates had a financial interest. They were also accused of steering recipients of Reading First grants away from other programs, including two (Reading Recovery and Success for All) that had been blessed as research-based by the federal What Works Clearinghouse. In most cases, individuals implicated in the scandal resigned from their federal posts to return previous positions or assume new ones before any were publicly asked to step down. A follow up query was initiated by House Education and Labor Committee (Manzo, 2007), but no further actions were taken.
5. The National Reading Panel (NRP) report was mandated directly by Congress and employed the relatively new approach of meta-analysis to distill from existing research what is known about the efficacy of teaching. Yet the selection of topics for review was much less systematic than one might think. The authors of the 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. At a meeting of the International Reading Association in 2006, S. Jay Samuels, one of the members of the NRP, announced that another criterion was at work in determining topics—the research interests of the panel members. This revelation suggests the strong possibility that some things did not get studied because no one on the panel found them compelling.
6. At the time we conducted this review, New York state’s Reading First evaluation, which would otherwise have been included in the review by these criteria, was not available. However, just prior to the time we submitted the chapter, a Powerpoint presentation of outcome data was posted on the New York Department of Education website. An examination of the data (without benefit of an interpretive narrative) convinced us that the inclusion of the New York data would not have altered our conclusions in this section.
7. The following peer-reviewed journals were reviewed: *American Educational Research Journal*,

*American Journal of Education, Educational Administration Quarterly, Educational Evaluation and Policy Analysis, Educational Policy Analysis Archives, Educational Policy, Elementary School Journal, Journal of Educational Administration, Journal of Educational Change, Journal of Education Policy, Journal of Literacy Research, Leadership and Policy in Schools, Peabody Journal of Education, Reading Research Quarterly, Reading Teacher, and Teachers College Record.*

8. That so many researchers and research organizations were found to be highly influential is actually unusual. For example, in his now classic analysis of the dynamics of policy making in health care and transportation, Kingdon (1984) found that while 66 percent of his respondents mentioned researchers or academics as players in these policy communities, only 15 percent of them rated them as very important.
9. See, also, Carnine 1999 for an insider account of policy making in California during the mid to late 1990s.
10. It is important to acknowledge that these are not always distinct policy initiatives. In fact, in many states, these policy approaches tended to be layered on top of one another. Thus, teachers tended to experience the cumulation of these policy strategies over time, rather than discrete policy initiatives.
11. Only 3 of the 5 state-level evaluation reports we reviewed actually reported on levels of implementation of Reading First.
12. Recent large-scale studies of in reading and other subject areas echo this finding about within school variability. For example, in their study of implementation of three Comprehensive School Reform models, Rowan and Correnti (2008) analyzed instructional logs of nearly 2000 teachers in 112 schools and found that 23 percent of the variance in use of instructional time was among teachers in a single school, while only five percent of the variance lies among schools. Similarly, in their study of teachers' response to standards-based accountability in mathematics in three states, Hamilton and her colleagues (2002) found that by far the greatest variability was between teachers within schools rather than between schools or between districts. The proportion of variance on most measures of classroom practice was 0.70 or higher within schools, compared to .20 or lower at the school or district level.
13. There is also some evidence of resistance in at least one large-scale evaluation of Reading First. DeStefano and colleagues (2006) report that respondents from 4 out of 18 Reading First sites that they interviewed over the phone mentioned that teacher resistance as one of the challenges they face. However, no further details are available.
14. Rowan et al. (2002) argues that researchers need a minimum of 15-20 classroom observations, spaced out across a school year, in order to draw valid conclusions about classroom instruction given the variability over time and the multi-faceted nature of classroom instruction.
15. In order to be included in this section, studies needed to make an explicit link to classroom practice. For example, a study on the role of principal in implementation of reading policy had to include data that linked principal actions to classroom instruction. Studies that focused on principal leadership—or other factors—absent that link were not included in this review.
16. This finding should be viewed with caution, however, because, as Stein and D'Amico (2002) note, they did not have longitudinal data for their study. Thus, they cannot say with certainty that teachers with implementation that is aligned and high quality moved on a developmental trajectory from superficial implementation to more substantive over time.
17. In California, the evaluators also created a "statistical control group" using statistical methods to illustrate how a school that is similar to Reading First schools would have performed without access to the program.
18. In Pennsylvania (which focused on third grade exclusively), Florida for the third grade FCAT, and California, the norming group for the state test was the entire state sample.

## REFERENCES

- Achinstein, B., & Ogawa, R. T. (2006). (In)fidelity: What the resistance of new teachers reveals about professional principles and prescriptive educational policies. *Harvard Educational Review*, 76(1), 30–63.



- Anagnostopoulos, D., & Rutledge, S. A. (2007). Making sense of school sanctioning policies in urban high schools. *Teachers College Record*, 109(5), 1261–1302.
- Allington, R. (1999). Critical issues: Crafting state educational policy: The slippery role of research and researchers. *Journal of Literacy Research*, 31(4), 457–482.
- Allington, R. L., & Woodside-Jiron, H. (1999). The politics of literacy teaching: How “research” shaped educational policy. *Educational Researcher*, 28(8), 4–12.
- Bean, R., Draper, J., Turner, G., & Zigmond, N. (in press). Reading first in Pennsylvania: Achievement findings after five years. *Journal of Literacy Research*.
- Birdyshaw, D., Pesko, E., Wixson, K., & Yochum, N. (2002). From policy to practice: Using literacy standards in early reading instruction. In M. L. Kamil, J. B. Manning, H. J. Walberg (Eds.) *Successful reading instruction* (pp. 75–99). Greenwich, CT: Information Age Publication/Laboratory for Student Success.
- Booher-Jennings, J. (2005). Below the bubble: “Educational triage” and the Texas accountability system. *American Educational Research Journal*, 42(2), 231–268.
- Burt, R.S. (1992). *Structural holes: The social structure of competition*. Cambridge, MA: Harvard University Press.
- California State Board of Education (1999). *Guide to the California reading initiative 1996–1999*. Sacramento: California Reading Initiative Center.
- Camilli, G., Vargas, S., & Yurecko, M. (2003). Teaching children to read: The fragile link between science and federal education policy [electronic version]. *Education Policy Analysis Archives*, 11(15).
- Camilli, G., Wolfe, P., & Smith, M. L. (2006). Meta-analysis and reading policy: Perspectives on teaching children to read. *The Elementary School Journal*, 107(1), 27–37.
- Carlson, C. D., Branum-Martin, L., Durand, A., Barr, C., & Francis, D. (2008). *Texas Reading First Initiative: Summary year 4 evaluation report*. University of Houston: Texas Institute for Measurement, Evaluation, and Statistics.
- Carnine, D. (1999). Campaigns for moving research into practice. *Remedial and Special Education*, 20(1), 2–6, 35.
- Coburn, C. E. (2001a). *Making sense of reading: Logics of reading in the institutional environment and the classroom*. Unpublished doctoral dissertation. Stanford University, Stanford, California.
- Coburn, C. E. (2001b). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145–170.
- Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education*, 77(3), 211–244.
- Coburn, C. E. (2005a). The role of nonsystem actors in the relationship between policy and practice: The case of reading instruction in California. *Educational Evaluation and Policy Analysis*, 27(1), 23–52.
- Coburn, C. E. (2005b). Shaping teacher sensemaking: School leaders and the enactment of reading policy. *Educational Policy*, 19(3), 476–509.
- Coburn, C. E. (2006). Framing the problem of reading instruction: Using frame analysis to uncover the microprocesses of policy implementation. *American Educational Research Journal*, 43(3), 343–379.
- Coburn, C. E., & Stein, M. K. (2006). Communities of practice theory and the role of teacher professional community in policy implementation. In M. I. Honig (Ed.), *Confronting complexity: Defining the field of education policy implementation* (pp. 25–45). Albany: State University of New York Press.
- Colvin, R. L. (1995, September 13). State report urges return to basics in teaching reading. *Los Angeles Times*, p. A1.
- Cross, C. T., Riley, R., Sanders, T. (2004). *Political education: National policy comes of age*. New York: Teachers College Press.
- Cusick, P. A., & Borman, J. (2002). Reform of and by the system: A case study of a state’s effort at curricular and systemic reform. *Teachers College Record*, 104(4), 765–786.
- Datnow, A., Borman, G., & Stringfield, S. (2000). School reform through a highly specified curriculum: Implementation and effects of the Core Knowledge Sequence. *The Elementary School Journal*, 101(2), 167–191.
- Datnow, A., & Castellano, M. (2000). Teachers’ responses to Success for All: How beliefs, experiences, and adaptations shape implementation. *American Educational Research Journal*, 37(3), 775–799.
- Dee, T., & Jacob, B. (2009). *The impact of No Child Left Behind on student achievement*. Cambridge, MA: National Bureau of Economic Research.

590 Cynthia E. Coburn et al.

- DeStefano, L., Hammer, V., & Fielder, E. (2006). Reading First external evaluation (No. 3). Urbana Campaign: University of Illinois at Urbana Champaign, College of Education.
- Diamond, J. B. (2007). Where the rubber meets the road: Rethinking the connection between high-stakes testing policy and classroom instruction. *Sociology of Education*, 80(4), 285–313.
- Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers College Record*, 106(6), 1145–1176.
- Dressman, M. (1999). On the use of misuse of research evidence: Decoding two states' reading initiatives. *Reading Research Quarterly*, 34(3), 258–285.
- Dutro, E., Fisk, M. C., Koch, R., Roop, L. J., & Wixson, K. K. (2002). When state policies meet local district contexts: Standards-based professional development as a means to individual agency and collective ownership. *Teachers College Record*, 104(4), 787–811.
- Edmondson, J. (2005, February 3). Policymaking in education: Understanding influences on the Reading Excellence Act. *Education Policy Analysis Archives*, 13(11). Retrieved December 28, 2006, from <http://epaa.asu.edu/epaa/v13n11/>
- Eisenhardt, M., & Towne, L. (2003). Contestation and change in national policy on “scientifically based” education research. *Educational Researcher*, 32(7), 31–38.
- Foorman, B. R., Petscher, Y., Lefsky, E., & Toste, J. R. (in press). Reading First in Florida: Five years of improvement. *Journal of Literacy Research*.
- Fuhrman, S. (2004). Introduction. In S. Fuhrman & R. Elmore (Eds.), *Redesigning accountability systems for education* (pp. 3–14). New York: Teachers College Press.
- Gandal, M. (1996). *Making standards matter 1996: An annual 50 state progress report on efforts to raise academic standards*. Washington, DC: American Federation of Teachers.
- Gamse, B. C., Jacob, R. ., Horst, M., Boulay, B., & Unlu, F. (2008). *Reading First impact study final report*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences. Washington, DC: U.S. Department of Education. (RFISFR) (NCEE 2009-4038).
- Gee, J. P. (1999). Reading and the new literacy studies: Reframing the National Academy of Sciences report on reading. *Journal of Literacy Research*, 31(3), 355–374. Retrieved February 9, 2008, from [http://findarticles.com/p/articles/mi\\_qa3785/is\\_199909/ai\\_n8857323/pg\\_7](http://findarticles.com/p/articles/mi_qa3785/is_199909/ai_n8857323/pg_7)
- Government Accounting Office. (2007). States report improvements in reading instruction, but additional procedures would clarify education's role in ensuring proper implementation. Washington, DC: Government Accounting Office.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78, 1360–1380.
- Granovetter, M. S. (1982). The strength of weak ties: A network theory revisited. In P. V. Marsden & N. Lin (Eds.), *Social structure and network analysis* (pp. 105–130). Beverly Hills, CA: Sage.
- Grossen, B. (1997). *Reading recovery: An evaluation of benefits and costs (executive summary)*. Eugene: University of Oregon.
- Guthrie, J. (1990). Educational evaluation and policy analysis, 12(3), 233–353.
- Haager, D., Dhar, R., Moulton, M., & McMillian, S. (2006). *The California Reading First Year 4 evaluation report*. Morgan Hill, CA: Educational Data Systems.
- Haager, D., Dhar, R., Moulton, M., McMillan, S. (2008). *The California Reading First Year 7 evaluation report*. Morgan Hill, CA: Educational Data Systems.
- Hamilton, L., Stecher, B., & Klein, S. (Eds.). (2002). *Making sense of test-based accountability in education*. Santa Monica, CA: RAND.
- Haney, W. (2000). The myth of the Texas Miracle in education [electronic version]. *Education Policy Analysis Archives*, 8(41). Retrieved from at <http://olam.ed.asu.edu/apaa>
- Heclo, H. (1978). Issue networks, and the executive establishment. In A. King (Ed.), *The New American political system* (pp. 87–124). Washington, DC: American Enterprise Institute.
- Hill, H.C. (2005). Content across communities: Validating measures of elementary mathematics instruction. *Educational Policy*, 19(3), 447–475.
- Hoffman, J. V., Assaf, L. C., & Paris, S. G. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher*, 54(5), 482–492.
- Hoffman, J. V., McCarthey, S. J., Elliot, B., Bayles, D. L., Price, D. P., Ferree, A., et al. (1998). The literature-based basals in first-grade classrooms: Savior, Satan, or same-old, same-old? *Reading Research Quarterly*, 33(2), 168–197.

- Jackson, R., McCoy, A., Pistorino, C., Wilkinson, A., Burghardt, J., Clark, M., et al. (2007). *National evaluation of early Reading First: Final report*. U.S. Department of Education, Institute of Education Sciences, Washington, DC: U.S. Government Printing Office.
- Jennings, N. E. (1996). *Interpreting policy in real classrooms: Case studies of state reform and teacher practice*. New York: Teachers College Press.
- Kauffman, D., Johnson, S. M., Kardos, S. M., Liu, E., & Peske, H. G. (2002). "Lost at sea": New teachers' experiences with curriculum and assessment. *Teachers College Record*, 104(2), 273–300.
- Kersten, J. (2006). Hybridization, resistance, and compliance: Negotiating policies to support literacy achievement. *The New Educator*, 2(2), 103–121.
- Kersten, J., & Pardo, L. (2007). Finessing and hybridizing: Innovative literacy practices in Reading First classrooms. *The Reading Teacher*, 61(2), 146–154.
- Kingdon, J. W. (1984). *Agendas, alternatives, and public policies*. Boston: Little, Brown.
- Levin, H. M. (1998). Accelerated schools: A decade of evolution. In A. Hargreaves, A. Lieberman, M. Fullan, & D. Hopkins (Eds.), *International handbook of educational change, part two* (pp. 807–830). Boston: Kluwer.
- Manzo, K. K. (2007, April 3). Reading probe will continue. Education Week. Retrieved from <http://www.edweek.org/ew/articles/2007/04/04/31read.h26.html>
- Massachusetts Department of Education. (2004). The Massachusetts Reading First newsletter: August 2004. Retrieved August 13, 2006, from <http://www.doe.mass.edu/read/newsletter/0905.doc>
- Massachusetts Department of Education. (2005a). Overview of the Massachusetts Reading First plan assessment framework. Retrieved July 24, 2006, from [http://www.doe.mass.edu/read/presentations/bsr\\_assessframework.pps](http://www.doe.mass.edu/read/presentations/bsr_assessframework.pps)
- Massachusetts Department of Education. (2005b). *Second visit monitoring tool for the Massachusetts Reading First Plan*. Retrieved July 21, 2006, from <http://www.doe.mass.edu/read/mrpf/visit2.doc>
- Matland, R. E. (1995). Synthesizing the implementation literature: The ambiguity-conflict model of policy implementation. *Journal of Public Administration and Theory*, 5(2), 145–174.
- Mazmanian, D., & Sabatier, P. A. (1983). *Implementation and public policy*. Glenview, IL: Scott, Foresman.
- McDaniel, J. E., Sims, C. H., & Miskel, C. G. (2001). The National Reading Policy arena: Policy actors and perceived influence. *Educational Policy*, 15(1), 92–114.
- McDonnell, L. M. (2004). *Politics, persuasion, and educational testing*. Cambridge, MA: Harvard University Press.
- McFarland, A.S. (1992). Interest groups and the policymaking process. In M. P. Petraccia (Ed.), *The Politics of interests: Interest groups transformed* (pp. 58–79) Boulder, CO: Westview Press.
- McGill-Franzen, A., Zmach, C., Solic, K., Zeig, J. (2002). The confluence of two policy mandates: Core reading programs and third-grade retention in Florida. *The Elementary School Journal*, 107(1), 67–93.
- McGill-Franzen, A., Ward, N., Goatley, V., & Machado, V. (2002). Teachers' use of new standards, frameworks, and assessments: Local cases of NYS elementary grade teachers. *Reading Research and Instruction*, 41(2), 127–148.
- Mesmer, H.A. & Karchmer, R.A. (2003). REALity: How the Reading Excellence act took form in two schools. *Reading Teacher*, 56(7), 636–645.
- MGT of America. (2008). *Illinois Reading First evaluation report: Final, 2007–08*. Washington, DC: author.
- Mintrop, H., & Trujillo, T. M. (2005). Corrective action in low performing schools: Lessons for NCLB implementation from first-generation accountability systems. *Education Policy Analysis Archives*, 13(48). Retrieved from <http://epaa.asu.edu/epaa/v13n48/>
- Miskel, C., & Song, M. (2004). Passing reading first: Prominence and processes in an elite policy network. *Educational Evaluation and Policy Analysis*, 26(2), 89–109.
- Moss, M., Fountain, A., Boulay, B., Horst, M., Rodger, C., & Brown-Lyons, M. (2008). *Reading First implementation evaluation: Final report*. Washington, DC: United States Department of Education.
- National Reading Panel Report. (2000). *Report of the National Reading Panel: Teaching children to read*. Washington, DC: National Institute for Literacy
- Ogawa, R. T., Sandholtz, J. H., & Scribner, S. P. (2004). Standards gap: Unintended consequences of local standards-based reform. *Teachers College Record*, 106(6), 1177–1202.

592 Cynthia E. Coburn et al.

- Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, 40(2), 184–202.
- Pearson, P. D. (2004). The reading wars. *Educational Policy*, 18(1), 216–252.
- Pressley, M., & Fingeret, L. (2007). *What we have learned since the National Reading Panel*. New York: Guilford Press.
- Pressman, J. L., & Wildavsky, A. (1984). *Implementation*. Berkeley: University of California Press.
- Reading First Advisory Committee (2008). *Response to the Reading First impact study*. Retrieved January 5, 2010, from <http://www.ed.gov/programs/readingfirst/statement.pdf>
- Remillard, J. T. (2000). Can curriculum materials support teachers' learning? Two fourth-grade teachers' use of a new mathematics text. *The Elementary School Journal*, 100(4), 331–350.
- Remillard, J. T. (2004). Teachers' orientations toward mathematics curriculum materials: Implications for teacher learning. *Journal for Research in Mathematics*, 35(5), 352–388.
- Rowan, B., Correnti, R., & Miller, R. (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the Prospects Study of elementary schools. *Teachers College Record*, 104(8), 1525–1567.
- Rowan, B., & Correnti, R. (2008). *Studying reading instruction with teacher logs: Lessons from a study of instructional improvement*. Unpublished manuscript. Ann Arbor: University of Michigan.
- Scott, C. (2007). *Reading First: Locally appreciated, nationally troubled*. Washington DC: Center on Educational Policy. Retrieved from <http://www.cep-dc.org/>
- Sloan, K. (2006). Teacher identity and agency in school worlds: Beyond and all-good/all-bad discourse on accountability-explicit curriculum policies. *Curriculum Inquiry*, 36(2), 119–152.
- Smith, M. S., & O'Day, J. (1990). Systemic school reform. *Journal of Education Policy*, 5(5), 233–267.
- Smith, M. S. (2000). Balancing old and new: An experienced middle school teacher's learning in the context of mathematics instructional reform. *The Elementary School Journal*, 100(4), 351–375.
- Snipes, J., Doolittle, F., & Herlihy, C. (2002). *Foundations for success: Case studies of how urban school systems improve student achievement*. DC: Council of Great City Schools.
- Snow, C. (2000). On the limits of reframing: Rereading the National Academy of Sciences report on reading. *Journal of Literacy Research*, 32(1), 113–120.
- Song, M., & Miskel, C. G. (2005). Who are the influencers? A cross-state social network analysis of the reading policy domain. *Educational Administration Quarterly*, 41(1), 7–48.
- Southgate, V. (1966). Approaching I. T. A. results with caution. *Reading Research Quarterly*, 1(3), 35–56.
- Spillane, J. P. (1999). External reform initiatives and teachers' efforts to reconstruct their practice: The mediating role of teachers' zones of enactment. *Journal of Curriculum Studies*, 31(2), 143–175.
- Spillane, J. P. (2000). Constructing ambitious pedagogy in the fifth grade: The mathematics and literacy divide. *The Elementary School Journal*, 100(4), 307–330.
- Spillane, J. P., & Jennings, N. E. (1997). Aligned instructional policy and ambitious pedagogy: Exploring instructional reform from the classroom perspective. *Teachers College Record*, 98(3), 449–481.
- Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, 72(3), 387–431.
- Stecher, B., Hamilton, L., & Gonzalez, G. (2003). *Working smarter to leave no child behind: Practical insights for school leaders*. Santa Monica, CA: RAND Education.
- Stein, M. K., & D'Amico, L. (2002). Inquiry at the crossroads of policy and learning: A study of a district-wide literacy initiative. *Teachers College Record*, 104(7), 1313–1344.
- Taylor, B. M., Anderson, R. C., Au, K. H., & Raphael, T. E. (2000). Discretion in the translation of research to policy: A case from beginning reading. *Educational Researcher*, 29(6), 16–26.
- Taylor, B. M., Pearson, P. D., Peterson, D. S., & Rodriguez, M. C. (2005). The CIERA school change framework: An evidence-based approach to professional development and school reading improvement. *Reading Research Quarterly*, 40(1), 2–32.
- Texas Education Agency. (2002). Student Success Initiative. Retrieved February 1, 2010, from <http://ritter.tea.state.tx.us/curriculum/ssi.html>
- Trochim, W. (1984). *Research design for program evaluation: The regression-discontinuity design*. Beverly Hills, CA: Sage
- United States Department of Education. (2002). *Guidance for the Reading First Program*. Retrieved Novem-

- ber 10, 2007, from <http://www.ed.gov/programs/readingfirst/guidance.pdf#search=%22reading%20first%20guidance%22>
- United States Department of Education. (2004). *Introduction and overview: Reading First*. Retrieved July 16, 2007, from <http://www.ed.gov/programs/readingfirst/index.html>
- Valencia, S. W., Place, N. A., Martin, S., D., & Grossman, P. L. (2006). Curriculum materials for elementary reading: Shackles and scaffolds for four beginning teachers. *The Elementary School Journal*, 107(1), 93–120.
- Valencia, S. W., & Wixson, K. K. (2000). Policy-oriented research on literacy standards and assessment. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 909–935). Mahwah, NJ: Erlbaum.
- Valencia, S.W., & Wixson, K. K. (2001). Inside English/Language Arts standards: What's in a grade? *Reading Research Quarterly*, 36(2), 202–211.
- Vasquez Heilig, J. & Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis*, 30(2), 75–110.
- Wright, W. E., & Choi, D. (2006). The impact of language and high-stakes testing policies on elementary school English language learners in Arizona. *Education Policy Analysis Archives*, 14(13), 1–75.
- Zigmond, N., & Bean, R. (2006). *External evaluation of Reading First in Pennsylvania: Annual report, project year 3: 2005–2006*. Unpublished manuscript.

Not for Distribution