

Alignment and Accountability in Policy Design and Implementation: The Common Core State Standards and Implementation Research

Cynthia E. Coburn¹, Heather C. Hill², and James P. Spillane¹

Both the Common Core State Standards (CCSS) and recent efforts to hold schools and teachers accountable have been hotly debated among practitioners, policymakers, and the public at large. Much of the debate centers on the merits and demerits of these initiatives and the general approach they represent to reforming teaching and learning. In this article, we focus on a different issue, that is, the opportunity to advance research on policy implementation afforded by the intertwined nature of CCSS and accountability efforts. Arguing that it is essential for stakeholders, regardless of their stance on either reform, to understand whether and how both influence classroom teaching and learning, we outline elements of a research agenda to generate knowledge important to the design of future instructional policies. For this to happen, we argue that an implementation research agenda needs to build on (rather than reinvent) lessons learned from the past quarter century of implementation scholarship on instructional policy. To that end, we review theoretical and empirical insights from implementation research on standards-based reform and outline specific avenues for potential theory testing research on educational policy implementation.

Keywords: accountability; educational policy; policy analysis; professional development

Policymakers are at it again, attempting to improve the quality of instruction in America's classrooms. One driver is the Common Core State Standards (CCSS), which put forth ambitious goals, for example, asking that students engage in disciplinary reasoning, develop the ability to build arguments and make inferences, and understand structure, similarities, and contrasts. Most agree that the CCSS require a significant departure from current practice in most U.S. classrooms. Despite political opposition and the loss of some states from the standards' assessment consortia, to date, more than 40 states have adopted CCSS. Another driver is rigorous accountability, including teacher evaluation systems that hold teachers and schools to specific standards for instruction and compare teachers' production of student outcomes to others in their school and district. Spurred in part by *Race to the Top*, most states in the country have implemented new accountability mechanisms, for example, revamping their teacher evaluation processes toward more exacting criteria and developing new strategies for school accountability such as turnaround schools.

Both of these policy strategies—rigorous new standards and new accountability mechanisms—seek to raise student achievement by influencing how teachers teach and how students learn.

This is not the first time the United States has tried these approaches. Standards-based reform efforts from the late 1980s through 2000s involved developing state standards with the goal of linking them to curriculum, professional development, and assessments (M. S. Smith & O'Day, 1991). The accountability movement began in the mid-1990s and intensified in the 2000s, creating tough rewards and sanctions for increasing student achievement. However, these earlier reforms met with mixed success. States and districts varied substantially in their ability to help teachers meet ambitious instructional goals (Cohen & Moffitt, 2010; McLaughlin, 1987; Spillane, 2004). Ambitious

¹Northwestern University, Evanston, IL

²Harvard University, Cambridge, MA

Authors are in alphabetical order by name.

standards were not well aligned to curricular materials, tests, and professional development; resources were not available to help teachers teach in new ways; and tests and accountability incentivized teachers to take only basic skills seriously (Coburn, Pearson, & Woulfin, 2011; Diamond, 2007). Test-based accountability led to some gains but also a range of unintended consequences like curriculum narrowing, increased test prep, and increased efforts to game the system (Diamond & Spillane, 2004; Firestone, Fitz, & Broadfoot, 1999; Firestone, Mayrowetz, & Fairman, 1998; Heilig & Darling-Hammond, 2008; Jacob, 2005; Jennings, 2010; McNeil, 2002; B. Smith, 1998; Valenzuela, 2004; Wilson & Floden, 2001).

But things have changed considerably in U.S. education policy in just a few decades. Many states have developed more sophisticated approaches to aligning learning standards in core subjects with student assessments, professional development, curriculum materials, and teacher evaluation (Fuhrman, Goertz, & Weinbaum, 2007; Martone & Sireci, 2009; Rentner, 2013). And states and districts are experimenting with new teacher and school accountability mechanisms that take into account a broader range of metrics than test scores alone (Doherty & Jacobs, 2013).

These developments in educational policy pose two challenges for educational research. First, we must understand *if* and *how* standards efforts such as the CCSS and new approaches to accountability *interact* to influence classroom teaching and learning.¹ Tracking the effects of the interaction of these policy initiatives is critically important to a broad constituency of policymakers, business leaders, academics, and other stakeholders. Second, researchers must build theory about the conditions that either encourage or discourage teacher learning and instructional change. Understanding the effects of and mechanisms around these new reform strategies is important not only to enable us to make midcourse corrections in current policies, but also because it promises to help policymakers design future policies that better support high-quality instruction in school districts, schools, and classrooms. These latest policy initiatives offer an opportunity to test hypotheses that have emerged from two decades of theory building work on education policy implementation.

In this article, we describe an approach for investigating the interaction of CCSS and new accountability mechanisms that builds on and extends lessons from the existing policy implementation literature. We begin by describing the historic and theoretical foundations for implementation research and then outline a research agenda that, we argue, can move the field of implementation research forward.

Three Eras in Instructional Policy and Policy Implementation Scholarship

Since policymakers turned their attention to instructional reform in the wake of *A Nation at Risk* (National Commission on Excellence in Education, 1983), there have been several efforts to establish ambitious learning goals for America's youth and several waves of research investigating their process and outcomes. Instructional policy refers to policies designed to influence some aspect of classroom instruction, including the content

teachers teach, acceptable levels of student mastery of this content, and/or specific pedagogies and teaching methods for content. We identify three eras of instructional policymaking and implementation research: the standards-based reforms of the 1980s and 1990s, accountability-based reforms of the late 1990s and 2000s, and today's efforts that combine aspects of both. While these are somewhat crude characterizations in that these policy eras overlapped and intermingled as policy initiatives tend to do in the United States, they serve to distinguish and thereby usefully situate the lines of research we propose here.

Standards Movement of the 1980s and 1990s

Starting in the late 1980s, professional associations and state policymakers worked to articulate more intellectually rigorous learning standards in core school subjects, including early efforts by states (e.g., California's curricular frameworks in mathematics and ELA in the 1980s), efforts by professional organizations (e.g., American Association for the Advancement of Science, 1993; joint standards from International Reading Association & National Council of Teachers of English, 1996; and National Council of Teachers of Mathematics, 1989), and the standards adopted by many states to secure Goals 2000 funding (Cross, 2004). These efforts were informed by a theory of systemic reform advanced by scholars such as Marshall Smith and Jennifer O'Day (1991), who argued for defining intellectually ambitious learning standards and then aligning other state instructional policy instruments (e.g., student assessment, teacher preparation and professional development, curricular materials) with these standards. The basic theory was that if policy provided more consistent instructional guidance around ambitious learning standards for local educators and aligned these standards to assessment, curriculum, and professional development, it would contribute to improvement in classroom teaching and learning (see also Porter, Floden, Freeman, Schmidt, & Schwillie, 1988).

Researchers studying implementation of standards-based reform found that the impact of standards on classroom practice was modest (Coburn, 2004; Cohen, 1990; Spillane & Zeuli, 1999; Weiss, Pasley, Smith, Banilower, & Heck, 2003). Teachers transformed innovative curricula into more traditional instruction (Wilson, 1990), selectively took up reform ideas (Coburn, 2004; Coburn et al., 2011; Wiemers, 1990), and adopted surface-level features (i.e., materials, student grouping arrangements) rather than making fundamental changes to their instruction, such as shifting classroom discourse patterns (Cohen, 1990; Spillane & Zeuli, 1999). Large-scale observational studies conducted in subsequent decades detected little residual evidence of the standards' successful transit from the statehouse to the schoolhouse and into the classroom (Hiebert et al., 2005; Kane & Staiger, 2010; Weiss et al., 2003).

A large body of scholarship explained these implementation patterns by pointing to teachers' and others' learning. Building on work by Berman and McLaughlin that highlighted the ways that policy is adapted at the local level (Berman & McLaughlin, 1978; McLaughlin, 1976), researchers in this era began to see this adaptation as a result of learning processes. Although scholars used different terms (e.g., interpretation, co-construction,

sense-making) and different theoretical frameworks (e.g., organizational theory, cognitive and sociocultural learning theories, social construction), this scholarship argued that implementers constructed understandings of policy in ways that influenced their response to policy. In this account, implementation varied because practitioners drew on prior knowledge and practices to interpret the reforms, leading them to construct policy messages in ways that either reinforced preexisting practices or focused on surface-level forms of the reform proposals (e.g., using story problems to teach mathematics as problem solving). This contributed to piecemeal and superficial changes in instructional practice (Coburn, 2001, 2004; Cohen, 1990; Firestone et al., 1998; Haug, 1999; Remillard, 1999; Shifter & Fosnot, 1993; M. S. Smith, 2000; Spillane, 1999, 2004; Spillane & Zeuli, 1999). These learning processes occurred not only among teachers but also among school leaders and state and district administrators charged with educating teachers about new standards, resulting in inconsistent and sometimes conflicting instructional guidance (Anagnostopolous & Rutledge, 2007; Spillane, 1996).

This line of research also argued that these patterns in learning and implementation resulted in part because features of the educational system led standards-based reform, ironically, to lack alignment in practice. There were few curriculum materials aligned with the standards (Cohen & Hill, 2001; Kardos, Johnson, Peske, Kauffman, & Liu, 2001; Kendall, 2011; Wilson, 2008). Tests were not always linked with new standards, and accountability systems incentivized teachers to take only basic skills seriously (Diamond, 2007; Hannaway & Hamilton, 2008; Shepard & Dougherty, 1991). Professional development, frequently a building block for educating teachers about standards, varied in quality and alignment (Cohen & Hill, 2001; Desimone, Porter, Birman, Garet, & Yoon, 2002; Hill, 2004; Little, 1993; Wilson & Berne, 1999). The multiple competing initiatives that existed in schools, for instance, democratic governance initiatives and often-numerous local programs, led to divergent demands on teachers and hindered the development of school-wide instructional coherence (Bryk & Rollow, 1992; Newmann, Smith, Allensworth, & Bryk, 2001). All of these conditions contributed to teachers and other local actors making their own sense of reform efforts as incoherent messages and weak infrastructure multiplied opportunities for divergent interpretations (Coburn & Woulfin, 2012; Kersten & Pardo, 2007; Spillane & Zeuli, 1999). As a result, classroom implementation often failed to reflect policymakers' intent and frequently looked different from one classroom to the next, even within the same school and district (Coburn, 2004; Spillane & Zeuli, 1999).

Accountability-Based Reforms of the 2000s

By the turn of the century, high-stakes accountability tied to student achievement on externally mandated tests was becoming a dominant feature in U.S. instructional policymaking. Although many states and local school districts had already implemented versions of test-based accountability prior to 2001 (Carnoy & Loeb, 2002; McDonnell, 2004), the No Child Left Behind (NCLB) Act (2001) increased the amount of testing, adding assessments in new grade levels and subject areas, such as science. States began to more aggressively publish results from testing ("Adequate Yearly Progress" reports) on the theory that public

pressure would serve as an important lever for accountability and published information on achievement gaps within schools and districts in hopes of decreasing those gaps. Local policymakers, pressed by consequences built into NCLB legislation, also mobilized much tougher sanctions, including closing low-performing schools and reopening them with new staff (e.g., school reconstitution in Chicago, Boston, San Francisco, and other cities).

Research on implementation of accountability policy provided evidence that policy did make its way inside local school systems to influence district, school, and classroom practice, though not always as intended and at times with detrimental effects. Policies tied to test-based high-stakes accountability influenced instruction and contributed to slight increases in student achievement (Au, 2007; Clotfelter & Ladd, 1996; Herman, 2004; Mintrop & Sunderman, 2009; Valli & Buese, 2007). However, there was tremendous variation among states and weak evidence that such testing narrowed the achievement gap (Jacob, 2005; Lee, 2007; Mintrop & Sunderman, 2009; Neal & Schanzenbach, 2010; Wong, Cook, & Steiner, 2009). In addition, these changes came with a price. Among other things, policy influenced what teachers taught but failed to improve how they taught it, marginalized low-stakes subjects, diverted resources to students based on their likelihood of passing the test, and increased the time devoted to teaching test-taking skills as distinct from the content being tested (Booher-Jennings, 2006; Diamond & Spillane, 2004; Firestone et al., 1998, 1999; Jacob, 2005; McNeil, 2002; B. Smith, 1998; Valenzuela, 2004; Wilson & Floden, 2001).

While a large segment of research during the standards era explained implementation patterns by focusing on learning, researchers studying implementation processes related to accountability turned to another explanation to explain these effects: power dynamics. This research emerged from multiple theoretical frameworks (e.g., organizational theory, critical theory, political science) that conceptualized power and the mechanism by which it influences practice in somewhat different ways. Some focused on what Lawrence (2008) calls episodic power, or "discrete, strategic acts of mobilization initiated by self-interested actors" (p. 172). These scholars have pointed to the ways that accountability policies attempt to *force* or *induce* changes in practice via mandates, incentives, and sanctions (e.g., Achinstein & Ogawa, 2006; Desimone, Smith, & Phillips, 2007; Hallett, 2010) or how those charged with implementing accountability policy—for example, school leaders and coaches—attempt to *influence* teachers to change their practice via normative pressure or appeals to expertise and other forms of informal authority (e.g., Coburn & Woulfin, 2012; Desimone et al., 2007; Woulfin, 2016). Other scholars have emphasized how force, inducements, and influence interact with systemic forms of power, or "ongoing practices to advantage particular groups without those groups necessarily establishing or maintaining those practices" (Lawrence, 2008, p. 172). These scholars argue that accountability policy interacts with institutionalized power relations in ways that maintain structures of inequality in schools (e.g., Au, 2007; Lipman, 2009; McNeil & Valenzuela, 2001). Taken together, this scholarship provides evidence that episodic and systemic uses of power led to compliance as teachers made changes in their practice as noted previously. But, policies that

relied on force and inducements led to considerable teacher resistance (Achinstein & Ogawa, 2006; Kersten, 2006; Kersten & Pardo, 2007; Sloan, 2006), symbolic or superficial responses (Hannaway & Hamilton, 2008; Sloan, 2006), less engagement with professional learning opportunities (Desimone et al., 2007), and an increase in gaming the system (Heilig & Darling-Hammond, 2008; Jennings, 2010).

In sum, there are at least two important practical findings from implementation research on these two eras of reform. First, state and national policy *can* influence school and classroom instructional practice, with effects most pronounced in tested subjects. Second, local implementation of state and national policy depends on alignment between those policies, district policies, and the educational infrastructure that has developed to support teachers' learning about and compliance with policy. These lessons have important implications for implementation theory, highlighting both learning processes and power dynamics as key mechanisms by which policy influences practice. Research has focused mostly on one explanation or the other, even in settings where both standards and accountability were in play.² And, we know little about how the design of accountability and standards policies influences these processes.

The Current Era: Alignment and Accountability

Current policy efforts offer an ideal opportunity to investigate the intersection of learning and power in policy implementation as states that have adopted the CCSS work to align standards with professional development and curriculum materials and introduce new accountability schemes. Teachers are experiencing CCSS and new accountability schemes concurrently as new tests are launched and curriculum materials and professional development are becoming available. Any effort to investigate one of these policy initiatives must take the presence of the other into account. The co-location of these policy strategies in the current era offers a unique opportunity to investigate how alignment and accountability interact and the consequences of that interaction for district, school, and classroom practice and ultimately for student learning outcomes. It is also an important opportunity to investigate the interaction of learning and power in implementation of instructional policy. Such work promises to reveal the consequences of policy design for student outcomes and provide insight into the mechanisms by which these outcomes are produced (or not) at multiple levels of the system.

Extending the Knowledge Base on Instructional Policy Implementation

Here, we propose a sample research agenda that illustrates how new studies can be designed to build on the existing research base and leverage this unique moment in the history of educational policymaking. Our intent is to propose a line of analysis that extends the knowledge base on implementation—one that both builds on prior implementation theory and takes seriously the interaction of policies and multiple mechanisms that may be at play. We argue that rather than learning and relearning the same lessons over and over again, as is common in this field, the next generation of implementation research will be most useful

if it is carefully structured to test hypotheses surfaced by prior studies and to illuminate processes that have yet to be explored systematically.

To accomplish this goal, researchers should take advantage of natural variation across states and districts to investigate how strength of accountability and degree of alignment influence the implementation of instructional policy. Although there is currently considerable policy intent to align elements of the instructional guidance infrastructure, history suggests that policymakers will face difficulties in this task (Achieve, 2008, 2011; Polikoff, 2012a, 2012b). It is likely that state and local policies will vary in the degree of alignment, that is, the degree to which standards, assessments, instructional materials, the focus of evaluation schemes, and professional development are coordinated with one another. It is also likely that they will vary in the degree to which they are coordinated horizontally (within any one level of government—national, state, or local) and/or vertically (across different levels of government) (Anderson, 2002; Polikoff & Porter, 2014). Similarly, we already know that there is substantial variability across states in the intended strength of new teacher evaluation systems (Herlihy et al., 2014) as well as the degree to which localities employ strong sanctions for underperforming schools. Thus, state and local policy systems are also likely to vary in the strength of the accountability, or the authority and power of the instructional guidance system. Variation in these two dimensions across states and districts creates an opportunity to systematically investigate how degree of alignment and strength of accountability interact in ways that foster or impede school change and instructional improvement.

We advocate focusing on these two variables for three reasons. First, states have wagered on alignment and accountability as two major design features in education policy, thus it seems important to investigate these key features of policy design. Second, prior scholarship has identified these factors as important to policy implementation but has not systematically tested how they interact with one another. Studying the interaction of these two dimensions thus provides an opportunity to move the field of policy implementation forward in significant ways. Third, studying level of alignment and strength of accountability provides a strategic opportunity to investigate how learning and power dynamics interact in the implementation process, thus providing a strong opportunity to build theory related to the mechanisms underlying implementation processes. Of course, other variables are crucial for implementation processes,³ and there are other policy frameworks that could be tested (e.g., ambiguity-conflict model, Matland, 1995; game theory, Hermans, Cunningham, & Slinger, 2014; policy attribute theory, Porter et al., 1988). The method we propose here could easily be adapted for studies of these as well.

In order to systematically test hypotheses that have emerged from prior generations of implementation scholarship, we argue for forecasting strong hypotheses regarding implementation based on prior research, then using theoretical sampling of districts or states to enable investigators to pattern-match empirical evidence against these hypotheses (e.g., King, Keohane, & Verba, 1994; Yin, 2003). Although such designs cannot make truly causal claims, the designs we propose here are the logical next step following decades of largely exploratory research on policy

Table 1
The Interaction of Accountability and Alignment

	Weak Accountability	Strong Accountability
Low alignment	1	2
High alignment	4	3

implementation. Theoretically sampled cases can also better facilitate theory building about the mechanism by which implementation outcomes are produced.

Based on implementation research, we have delineated a set of hypotheses regarding local responses to various combinations of alignment and accountability (see Table 1).

Weak Accountability/Low Alignment

We hypothesize that we will see little change in instructional practice under conditions of weak accountability and low alignment (Quadrant 1). Teachers who seek to make changes in practice will experience mixed messages that work against sustained focus (Porter et al., 1988) and a lack of support to strengthen their enactments (Cohen & Hill, 2001), which may lead to hybrid practice (where new approaches are layered on top of existing ones) or superficial enactment. Those who do not seek to change their practice will not have consistent or focused pressure to do so.

Strong Accountability/Low Alignment

We know from research that the force and inducement found in strong accountability systems are likely to encourage some degree of classroom change but that change may be superficial (i.e., affect content rather than pedagogy, increase test preparation activities) and accompanied by increased resistance and efforts to game the system (Achinstein & Ogawa, 2006; Desimone, 2002; Heilig & Darling-Hammond, 2008; Kersten, 2006; Kersten & Pardo, 2007; Sloan, 2006). We also know that low alignment creates multiple and conflicting messages to teachers, often leading to limited or superficial response (Coburn, 2004). Thus, we hypothesize that in systems with strong accountability but low alignment (Quadrant 2 in Table 1), teachers will have high levels of resistance, gaming, symbolic, and/or partial and superficial implementation. Learning opportunities are more likely to be fragmented and weak such that teachers may not even understand much about what they are resisting or how to implement in substantive ways when given the opportunity.

Strong Accountability/High Alignment

We suspect that the resistance put forth by teachers in states and districts with strong accountability systems may be eased by a system with high alignment within the instructional guidance system and attendant support for teacher learning (Quadrant 3 in Table 1). Teachers would have the opportunity to develop a more comprehensive understanding of new instructional approaches promoted by CCSS because they would be

supported in their enactment by multiple levers: materials, assessments, and professional development (Newmann et al., 2001). Strong accountability would focus attention on new instructional approaches being promoted by CCSS. Thus, in systems with strong accountability and alignment, there may be more substantive implementation of CCSS instructional practices.

Weak Accountability/High Alignment

We hypothesize that in systems with weak accountability but high alignment within the instructional guidance system (Quadrant 4 in Table 1), there is likely to be less resistance but more uneven implementation. Teachers who embrace CCSS have opportunities and support to deepen their enactment (Coburn, 2004), but there are fewer incentives for those not already inclined to shift their practice to do so, resulting in limited or superficial implementation among a segment of teachers.

These hypotheses, while drawn from existing research, require careful and systematic testing. The most obvious method would be systematic cross-case analyses of states and districts that vary in their degree of alignment of instructional guidance and strength of accountability systems. Scholars could strategically sample states or districts from the aforementioned typology and investigate how various combinations of these two dimensions affect classroom instruction. Longitudinal designs are likely necessary because we know it takes a long time for policies to move through the system and into schools but also because longitudinal analyses are well suited for investigating process, in this case the process by which these policy strategies shape classroom practice, if at all. Such longitudinal cross-case analyses can enable theory elaboration, critical should original hypotheses not be borne out (Vaughan, 1992) and essential for identifying unintended consequences.

We argue that we also need quasi-experiments to test our hypotheses quantitatively. Some researchers have used quasi-experiments and longitudinal analyses to study prior policy initiatives (e.g., Carnoy & Loeb, 2002; Dee & Jacob, 2009), and similar work could be organized around the CCSS and current accountability efforts. Designs of this type would benefit by being meaningfully joined to information about what states and districts are doing on the ground. For instance, surveys of district responses to different levels of alignment and accountability could be used to explain state- and district-level trends in student outcomes over time. Classroom observation work supplemented by interviews could similarly be linked to performance trends within districts.

As noted previously, this typology and proposed method is simply a template. Scholars could design studies using this approach that focus on other variables that promise to inform current policy debates and build on lessons from implementation research. Study design features and the scope of outcome data collection would likely be the same, although scholars would need to customize both to the unique predictions of theory. Rigorously testing theory about the conditions that impede or accelerate instructional policy implementation might occur across several studies simultaneously.

Conclusion

For a quarter-century, theories of action related to alignment and accountability have figured prominently in policy discourses in the United States. While stakeholders debate the merits and demerits of specific initiatives like the CCSS, alignment and accountability have, over the past 25 years, become fixtures of U.S. education policy. Over the same time period, scholars studying these efforts have generated considerable empirical knowledge about relations between policy design and policy implementation for each movement mostly separately. The co-emergence of CCSS and new accountability strategies and the variation among states and districts that will surely exist in the coming years offers an opportunity to extend the practical and theoretical knowledge base about policy design and implementation.

Practically, a systematic research agenda such as this one could help policymakers gauge the extent of implementation and influence of CCSS and accountability policies. Lessons from this research could also be distilled into a set of practices for policymakers and administrators working to design systems for the improvement of classroom learning. Theoretically, the studies can help us understand the intersection of power and learning. To date, scholars of implementation have tended to focus on either the dynamics of learning or relations of power, but a more robust understanding of the process and outcomes of implementation requires that we understand the two in interaction.

Of course, several other dimensions of policy design and the implementation process likely matter for implementation, including: system-level, organizational, and individual capacity; organizational networks and environments; the specificity of policy; and the ambitiousness of the instructional ideas advanced by policy. Other research programs might foreground combinations of these variables. However, we argue that accountability and alignment constitute such powerful contexts in U.S. public schools right now that it is likely important to attend to how these additional variables function within settings that vary systematically by strength of accountability system and level of alignment.

Organizing such a program of research will likely be difficult. Research typically lags behind policy itself and often even lags behind policy implementation, as shown by the 2015 funding of the Center for Standards in Schools five years after the passage of the CCSS and well into the implementation period in many districts. Full implementation of ambitious instructional policy often has fuzzy boundaries and typically takes many years to accomplish as publishers, professional developers, districts, and schools orient themselves and then build capacity to support new forms of instruction. Policies tend to shift over time in response to controversies and push-back from constituents. All of this implies prospective, longitudinal research as well as reliance on historical documentation, such as the NSF-funded multiyear database on state policies and policy instruments now under construction at NORC (NORC, 2015). It also may require funding agencies committed to understanding policy implementation to provide grants to study implementation “in the wild” rather than as a mediator variable within cluster-randomized trials.⁴

Given the current state of the research literature in implementation, now is the time for more targeted studies of key

high-leverage dimensions identified by prior research on implementation. We have learned much from earlier generations of implementation scholarship. We believe that one core component of the next generation of implementation studies is research designs that enable investigation of a few key dimensions linked to current policy initiatives as a way to extend the knowledge base and inform the practice of policymaking and implementation in actionable ways.

NOTES

¹These questions will also apply to the Next Generation Science Standards (NGSS) when they become fully implemented in the 15 states that have adopted them. Other states have not adopted NGSS, but many have rewritten their standards in ways that draw on this approach. Teachers and schools in these states will face similar challenges to those in NGSS states.

²Desimone (2002) is an exception.

³Prior implementation research has identified other variables as important, including system-level, organizational, and individual-level capacity (Chrispeels, 1997; Honig, 2003); organizational field and network strength (e.g., Daly & Finnigan, 2011; Frank, Zhao, Penuel, Ellefson, & Porter, 2011; Penuel, Frank, & Krause, 2010); and the degree of ambitiousness or specificity of policy itself (e.g., Mazmanian & Sabatier, 1983).

⁴Thanks to an anonymous reviewer for raising some of these issues.

REFERENCES

- Achieve, Inc. (2008). *Closing the expectations gap 2008: An annual 50-state progress report on the alignment of high school policies with the demands of college and careers*. Washington, DC: Author.
- Achieve, Inc. (2011). *Closing the expectations gap 2011: Sixth annual 50-state progress report on the alignment of high school policies with the demands of college and careers*. Washington, DC: Author.
- 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.
- American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York, NY: Oxford University Press.
- Anagnostopoulos, D., & Rutledge, S. A. (2007). Making sense of school sanctioning policies in urban high schools: Charting the depth and drift of school and classroom change. *Teachers College Record, 109*(5), 1261–1302.
- Anderson, L. (2002). Curricular alignment: A re-examination. *Theory into Practice, 41*(4), 255–260.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258–267.
- Berman, P., & McLaughlin, M. W. (1978). *Federal programs supporting educational change, Vol. VIII: Implementing and sustaining innovations*. Santa Monica, CA: RAND.
- Booher-Jennings, J. (2006). Rationing education in an era of accountability. *Phi Delta Kappan, 87*(10), 756–761.
- Bryk, A. S., & Rollow, S. G. (1992). The Chicago experiment: Enhanced democratic participation as a lever for school improvement. *Issues in Restructuring Schools, 3*(Fall), 3–8.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis, 24*(4), 305–331.
- Chrispeels, J. H. (1997). Educational policy implementation in a shifting political climate: The California experience. *American Educational Research Journal, 34*(3), 453–481.

- Clotfelter, C. T., & Ladd, H. F. (1996). Recognizing and rewarding success in public schools. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 23–64). Washington, DC: The Brookings Institution.
- Coburn, C. E. (2001). 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., Pearson, P. D., & Woulfin, S. L. (2011). Reading policy in the era of accountability. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research* (Vol. 4, pp. 629–653). New York, NY: Routledge.
- Coburn, C. E., & Woulfin, S. L. (2012). Reading coaches and the relationship between policy and practice. *Reading Research Quarterly*, 47(1), 5–30.
- Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis*, 12(3), 311–329.
- Cohen, D. K., & Hill, H. C. (2001). *Learning policy: When state education reform works*. New Haven, CT: Yale University Press.
- Cohen, D. K., & Moffitt, S. L. (2010). *The ordeal of equality: Did federal regulation fix the schools?* Cambridge, MA: Harvard University Press.
- Cross, C. (2004). *Political education: National policy comes of age*. New York, NY: Teachers College Press.
- Daly, A. J., & Finnigan, K. S. (2011). The ebb and flow of social network ties between district leaders under high-stakes accountability. *American Educational Research Journal*, 48(1), 39–79.
- Dee, T., & Jacob, B. (2009). *The impact of No Child Left Behind on student achievement*. Cambridge, MA: National Bureau of Economic Research.
- Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research*, 72(3), 433–479.
- Desimone, L., Porter, A., Birman, B., Garet, M., & Yoon, K. S. (2002). How do district management and implementation strategies relate to the quality of professional development that districts provide to teachers? *Teachers College Record*, 104(7), 1265–1312.
- Desimone, L. M., Smith, T. M., & Phillips, K. J. R. (2007). Does policy influence mathematics and science teachers' participation in professional development? *Teachers College Record*, 109(5), 1086–1122.
- 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.
- Doherty, K. M., & Jacobs, S. (2013). *State of the states 2013: Connect the dots: Using evaluations of teacher effectiveness to inform policy and practice*. Retrieved from http://www.nctq.org/dmsStage/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report
- Firestone, W. A., Fitz, J., & Broadfoot, P. (1999). Power, learning, and legitimation: Assessment implementation across levels in the United States and United Kingdom. *American Educational Research Journal*, 36(4), 759–593.
- Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis*, 20(2), 95–113.
- Frank, K. A., Zhao, Y., Penuel, W. R., Ellefson, N., & Porter, S. (2011). Focus, fiddle, and friends: Experiences that transform knowledge for the implementation of innovations. *Sociology of Education*, 84, 137–156.
- Fuhrman, S. H., Goertz, M. E., & Weinbaum, E. H. (2007). Educational governance in the United States: Where are we? How did we get there? Why should we care? In S. Fuhrman, D. Cohen, & F. Mosher (Eds.), *The state of education policy research* (pp. 41–61). Mahwah, NJ: Erlbaum.
- Hallett, T. (2010). The myth incarnate: Recoupling processes, turmoil, and inhabited institutions in an urban elementary school. *American Sociological Review*, 75(1), 52–74.
- Hannaway, J., & Hamilton, L. (2008). *Performance-based accountability policies: Implications for school and classroom practices*. Washington, DC: Urban Institute and RAND Corporation.
- Haug, C. A. (1999). *Local understanding, resources, and policies: Obstacles to standards-based mathematics education reform* (Doctoral dissertation). University of Colorado.
- Heilig, J. V., & 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.
- Herlihy, C., Karger, E., Hill, H. C., Kraft, M. A., Williams, M., & Howard, S. (2014). State and local efforts to investigate the validity and reliability of scores from teacher evaluation systems. *Teachers College Record*, 116(1), 1–28.
- Herman, J. L. (2004). The effects of testing on instruction. In S. H. Fuhrman & R. F. Elmore (Eds.), *Redesigning accountability systems for education* (pp. 141–166). New York, NY: Teachers College Press.
- Hermans, L., Cunningham, S., & Slinger, J. (2014). The usefulness of game theory as a method for policy evaluation. *Evaluation*, 20, 10–25.
- Hiebert, J., Stigler, J. W., Jacobs, J. K., Givvin, K. B., Garnier, H. E., Smith, M. S., . . . Gallimore, R. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 video study. *Educational Evaluation and Policy Analysis*, 27(2), 111–132.
- Hill, H. C. (2004). Professional development standards and practices in elementary school mathematics. *Elementary School Journal*, 104(3), 215–231.
- Honig, M. I. (2003). Building policy from practice: District central office administrators' roles and capacity for implementing collaborative education policy. *Educational Administration Quarterly*, 39(3), 292–338.
- International Reading Association & National Council of Teachers of English. (1996). *Standards for the English language arts*. Newark, DE: Author.
- Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, 89(5–6), 761–796.
- Jennings, J. L. (2010). School choice or schools' choice? Managing in an era of accountability. *Sociology of Education*, 83(3), 227–247.
- Kane, T. J., & Staiger, D. O. (2010). *Learning about teaching: Initial findings from the Measures of Effective Teaching Project*. Seattle, WA: Bill & Melinda Gates Foundation.
- Kardos, S. M., Johnson, S. M., Peske, H. G., Kauffman, D., & Liu, E. (2001). Counting on colleagues: New teachers encounter the professional cultures of their schools. *Educational Administration Quarterly*, 37(2), 250–290.
- Kendall, J. S. (2011). *Understanding Common Core State Standards*. Alexandria, VA: ASCD.

- 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.
- King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.
- Lawrence, T. B. (2008). Power, institutions, and organizations. In R. Greenwood, C. Oliver, K. Sahlin, & R. Suddaby (Eds.), *Handbook of organizational studies* (pp. 215–254). London: Sage.
- Lee, C. D. (2007). *Culture, literacy, and learning: Taking bloom in the midst of the whirlwind*. New York, NY: Teachers College Press.
- Lipman, P. (2009). “Politics by other means”: Education accountability and the surveillance state. In T. Monahan & R. D. Torres (Eds.), *Schools under surveillance: Cultures of control in public education*. Rutgers, NJ: Rutgers University Press.
- Little, J. W. (1993). Teachers’ professional community in a climate of educational reform. *Educational Evaluation and Policy Analysis*, 15(2), 129–151.
- Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research*, 79(4), 1332–1361.
- 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. A., & Sabatier, P. A. (1983). *Implementation and public policy*. Lanham, MD: University Press of America.
- McDonnell, L. M. (2004). *Politics, persuasion, and educational testing*. Cambridge, MA: Harvard University Press.
- McLaughlin, M. (1976). Implementation as mutual adaptation: Change in classroom organization. *The Teachers College Record*, 77(3), 339–351.
- McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis*, 9(2), 171–178.
- McNeil, L. M. (2002). Private asset or public good: Education and democracy at the crossroads. *American Educational Research Journal*, 39(2), 243–248.
- McNeil, L., & Valenzuela, A. (2001). The harmful impact of the TAAS system of testing on Texas: Beneath the accountability rhetoric. In M. Kornhaber & G. Orfield (Eds.), *Raising standards or raising barriers: Inequality and high stakes testing in public education* (pp. 127–150). New York, NY: Ford Foundation.
- Mintrop, H., & Sunderman, G. L. (2009). Predictable failure of federal sanctions-driven accountability from school improvement—And why we may retain it anyway. *Educational Researcher*, 38(5), 353–364.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: Author.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- Neal, D. A., & Schanzenbach, D. (2010). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics*, 92(2), 263–283.
- Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001). *School instructional program coherence: Benefits and challenges*. Chicago, IL: Consortium on Chicago School Research.
- No Child Left Behind Act of 2001, 20 U.S.C. § 9101 C.F.R. (2001).
- NORC. (2015). *Developing a system for tracking state assessment policies in science and mathematics education: Study design paper*. Chicago, IL: Author.
- Penuel, W. R., Frank, K. A., & Krause, A. (2010). Between leaders and teachers: Using social network analysis to examine the effects of distributed leadership. In A. J. Daly (Ed.), *Social network theory and educational change*. Cambridge, MA: Harvard Education Press.
- Polikoff, M. S. (2012a). The association of state policy attributes with teachers’ instructional alignment. *Educational Evaluation and Policy Analysis*, 34(3), 278–294.
- Polikoff, M. S. (2012b). Instructional alignment under No Child Left Behind. *American Journal of Education*, 118(3), 341–638.
- Polikoff, M. S., & Porter, A. C. (2014). Instructional alignment as a measure of teaching quality. *Educational Evaluation and Policy Analysis*, 36(4), 399–416.
- Porter, A. C., Floden, R., Freeman, D., Schmidt, W., & Schwillie, J. (1988). Content determinants in elementary school mathematics. In D. A. Grouws & T. J. Cooney (Eds.), *Perspectives on research on effective mathematical teaching* (pp. 96–113). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Remillard, J. T. (1999). Curriculum materials in mathematics education reform: A framework for examining teachers’ curriculum development. *Curriculum Inquiry*, 29(3), 315–342.
- Rentner, D.S. (2013). *Year 3 of implementing the Common Core State Standards: An overview of states’ progress and challenges*. Washington, DC: Center for Education Policy.
- Shepard, L. A., & Dougherty, K. C. (1991). *Effects of high-stakes testing on instruction*. (ERIC No. ED337468)
- Shifter, D., & Fosnot, C. T. (1993). *Reconstructing mathematics education: Stories of teachers meeting the challenge of reform*. New York, NY: Teachers College Press.
- Sloan, K. (2006). Teacher identity and agency in school worlds: Beyond the all-good/all-bad discourse on accountability-explicit curriculum policies. *Curriculum Inquiry*, 36(2), 119–152.
- Smith, B. (1998). *It’s about time: Opportunities to learn in Chicago’s elementary schools*. Chicago, IL: Consortium on Chicago School Research.
- 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.
- Smith, M. S., & O’Day, J. A. (1991). Systemic school reform. In S. H. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing: The 1990 yearbook of the Politics of Education Association* (pp. 233–267). New York, NY: Falmer Press.
- Spillane, J. P. (1996). School districts matter: Local educational authorities and state instructional policy. *Educational Policy*, 10(1), 63–87.
- 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), 143–176.
- Spillane, J. P. (2004). *Standards deviation: How local schools misunderstand policy*. Cambridge, MA: Harvard University Press.
- Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis*, 21(1), 1–27.
- Valenzuela, A. (Ed.). (2004). *Leaving children behind: How “Texas-style” accountability fails Latino youth*. Albany, NY: SUNY Press.
- Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal*, 44(3), 519–558.
- Vaughan, D. (1992). Theory elaboration: The heuristics of case analysis. In C. C. Ragin & H. S. Becker (Eds.), *What is a case?: Exploring the foundations of social inquiry*. Cambridge, UK: Cambridge University Press.
- Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *Looking inside the classroom: A study of K–12 math-*

- ematics and science education in the United States*. Chapel Hill, NC: Horizon Research.
- Wiemers, N. J. (1990). Transformation and accommodation: A case study of Joe Scott. *Educational Evaluation and Policy Analysis*, 12(3), 281–292.
- Wilson, S. M. (2008). *California dreaming: Reforming mathematics education*. New Haven, CT: Yale University Press.
- Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education*, 24, 173–209.
- Wilson, S. M., & Floden, R. E. (2001). Hedging bets: Standards-based reform in classrooms. In S. H. Fuhrman (Ed.), *From the Capitol to the classroom: Standards-based reform in the States* (pp. 193–216). Chicago, IL: University of Chicago Press.
- Wong, M., Cook, T. D., & Steiner, P. M. (2009). *Comparison groups in short interrupted time-series: An illustration evaluating No Child Left Behind*. Evanston, IL: Society for Research on Educational Effectiveness.
- Woulfin, S. L. (2016). Duet or duel: A portrait of two logics of reading instruction in an urban school district. *American Journal of Education*. Advance online publication. doi:10.1086/685848.
- Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: SAGE.

AUTHORS

CYNTHIA E. COBURN, PhD, is a professor at the School of Education and Social Policy at Northwestern University, 2120 Campus Drive, Evanston, IL 60208; cynthia.coburn@northwestern.edu. Her research focuses on the relationship between instructional policy and teachers' classroom practice in urban schools, the relationship between research and policymaking, and scale-up of innovative programs.

HEATHER C. HILL is the Jerome T. Murphy Professor of Education at the Harvard Graduate School of Education, 6 Appian Way, Cambridge, MA 02138; heather_hill@harvard.edu. Her research focuses on teacher and teaching quality.

JAMES P. SPILLANE, PhD, is the Spencer T. and Ann W. Olin Professor in Learning and Organizational Change at the School of Education and Social Policy at Northwestern University, 2120 Campus Drive, Evanston, IL 60208; j-spillane@northwestern.edu. His research focuses on issues of education policy, policy implementation, school reform, and school leadership.

Manuscript received September 25, 2014
 Revisions received December 3, 2015, and April 22, 2016
 Accepted May 1, 2016