Rethinking Scale: Moving Beyond Numbers to Deep and Lasting Change

by Cynthia E. Coburn

The issue of "scale" is a key challenge for school reform, yet it remains undertheorized in the literature. Definitions of scale have traditionally restricted its scope, focusing on the expanding number of schools reached by a reform. Such definitions mask the complex challenges of reaching out broadly while simultaneously cultivating the depth of change necessary to support and sustain consequential change. This article draws on a review of theoretical and empirical literature on scale, relevant research on reform implementation, and original research to synthesize and articulate a more multidimensional conceptualization. I develop a conception of scale that has four interrelated dimensions: depth, sustainability, spread, and shift in reform ownership. I then suggest implications of this conceptualization for reform strategy and research design.

After many years of intense educational reform, educators, policymakers, and researchers still grapple with the question of how pockets of successful reform efforts might be "scaled up." This issue has attained an increasingly high profile in the United States with the rise of prominent reform networks (e.g., Accelerated Schools, Coalition of Essential Schools [CES], Comer Schools, Success for All, and the New American Schools projects) and increased federal funding for research-based comprehensive school change initiatives. Yet as the issue of "scale" emerges as one of the key challenges for educational reform, it remains largely undertheorized in the educational literature (Elmore, 1996; Gamson, 1998).

To date, most educational research that focuses on scale has tended to define it in unidimensional ways, involving solely or predominantly the expansion of numbers of schools reached by a given reform effort. But taking an external reform initiative to scale is a complex endeavor. It not only involves spreading reform to multiple teachers, schools, and districts as highlighted by conventional definitions, it also involves all the challenges of implementing reform documented by decades of implementation research (Elmore, 1996) and of sustaining change in a multilevel system characterized by multiple and shifting priorities (McLaughlin & Mitra, 2001; Stokes, Sato, McLaughlin, & Talbert, 1997). It is the simultaneity of these challenges, in all their complexity, that makes the problem of scale fundamentally multidimensional. While there is a small but growing body of work that raises theoretical challenges to the predominant definition and provides evidence for the multidimensional nature of scale, this work has yet to be brought together and synthesized. Thus, the traditional definition continues to hold considerable weight, framing most empirical studies and forming the foundation of many theoretical discussions on scale.

How educational researchers and reformers define scale matters, for it influences both the ways reformers and policymakers craft reform strategies and the ways researchers study the problem of scale. As Hatch (1998) argues, reformers draw on sets of assumptions—both explicit and unarticulated—about the goals, challenges, and processes of change as they develop strategies for working with schools and districts. Notions of the nature of scale constitute one set of these assumptions and, as such, shape the kinds of choices reformers make.[1] And for researchers, different definitions of scale focus the analytic eye in strategically different ways, suggesting alternative indicators of the processes and outcomes of scale.

In this article, I bring together seeds of an alternative conceptualization from literature on scale with relevant research from reform implementation and my own research to synthesize and articulate an elaborated conceptualization of scale. I argue that definitions of scale must include attention to the nature of change in classroom instruction; issues of sustainability; spread of norms, principles, and beliefs; and a shift in ownership such that a reform can become self-generative. In so doing, I hope to address key issues that research suggests are central to the challenges of implementing and sustaining external reform initiatives in multiple classrooms, schools, and districts. I then suggest implications of this elaborated conception for both research and reform strategy, arguing that it calls for researchers to broaden research designs to capture heretofore neglected outcomes, and for reformers to direct increased attention to additional dimensions of scale that may prove critical to schools' abilities to sustain and deepen reform over time. Finally, I highlight tensions raised by this conception, tensions that I believe stem from the very multidimensionality of the construct.

Reconceptualizing Scale

Most research on scale tends to define what it means to "scale up" an external reform in quantitative terms, focusing on increasing the number of teachers, schools, or districts involved (Blum, 1997; Bodily, 1998; Datnow, Hubbard, & Mehan, 2002; Datnow, Stringfield, McHugh, & Hacker, 1998; Fullan, 2000; Hargreaves & Fink, 2000; Honig, 1994; Hubbard & Mehan, 1999; Klein, McArthur, & Stecher, 1995; Legters, Balfanz, Jordan, & McPartland, 2002; McDermott, 2000; Slavin, 1997; Slavin & Fashola, 1998; Slavin & Madden, 1994; Smith et al., 1998;

Stringfield & Datnow, 1998; Stringfield, Datnow, Ross, & Snively, 1998). In an admirably concise formulation of the predominant view, Stringfield and Datnow define scaling up as "the deliberate expansion to many settings of an externally developed school restructuring design that previously has been used successfully in one or a small number of school settings" (1998, p. 271). Within this conceptualization, theorists differ in whether they define scale to involve *replication* of the reform in greater numbers of schools (Cooper, Slavin, & Madden, 1997; Fuchs & Fuchs, 1998; Legters et al.; Slavin; Slavin & Madden; Taylor, Nelson, & Adelman, 1999) or emphasize a process of *mutual adaptation* (Datnow et al., 2002; Hubbard & Mehan; Klein et al.; Mead & Simon, 1996; Stringfield & Datnow) whereby schools are encouraged to adapt reform models to the needs of their local context. Another variation of this theme incorporates concerns for *geographic proximity*, defining scale in terms of an increase in the number of schools involved in a reform effort to achieve a critical mass in a bounded area such as a school district (Bodilly). But whether by replication or adaptation, within a bounded geographic area or not, at root, for these theorists, the problem of scale tends to be framed, at least explicitly, as the problem of increasing the numbers of teachers, schools, or districts involved in a reform.

This definition is attractive in its simplicity, its intuitiveness, and its measurability. But what does it really mean to say that a reform program is scaled up in these terms? It says nothing about the nature of the change envisioned or enacted or the degree to which it is sustained, or the degree to which schools and teachers have the knowledge and authority to continue to grow the reform over time. By focusing on numbers alone, traditional definitions of scale often neglect these and other qualitative measures that may be fundamental to the ability of schools to engage with a reform effort in ways that make a difference for teaching and learning.

There is a growing body of work, however, that raises questions about traditional definitions of scale, suggesting, among other things, the need for greater attention to the depth of implementation and a shift in reform ownership (e.g., Elmore, 1996; McLaughlin & Mitra, 2001). Other studies, many of which explicitly invoke the unidimensional definition of scale, nevertheless provide evidence for its multidimensional nature (e.g., studies that highlight the challenges of sustainability or suggest elements required for a shift in knowledge and authority). Finally, there are a few studies that provide initial guidance about the kinds of research designs that may be necessary to capture additional dimensions of scale (e.g., Corcoran, 2003; Datnow, Borman, & Stringfield, 2000; Datnow & Yonezawa, in press). Taken together, the results of these various research streams suggest that an expanded and refined definition of scale is in order.

In this article, I draw on a comprehensive review of existing theoretical and empirical literature on scale,[2] relevant research on reform implementation, and our own research on attempts to take an external reform called the Child Development Project (CDP) to scale[3] to synthesize and articulate an elaborated conceptualization of the problem of scale. I argue that expanding a reform to multiple settings is a necessary but insufficient condition for scale. That is, scaling up not only requires spread to additional sites, but also consequential change in classrooms, endurance over time,

and a shift such that knowledge and authority for the reform is transferred from external organization to teachers, schools, and districts. Thus, I propose a conceptualization of scale comprised of four interrelated dimensions: depth, sustainability, spread, and shift in reform ownership. In the sections that follow, I take each of these four dimensions in turn, developing and illustrating the concept. I then suggest implications for reform strategy and research design.[4]

Depth

In practice, most reformers and researchers place a high priority on the nature and quality of implementation of reforms that are being taken to scale. Yet this priority is not embedded in traditional definitions of scale, the numerical emphasis of which often obscures questions of what counts as school change. At best, the absence of explicit attention to the nature and quality of change provides limited conceptual guidance on what it really means for an external reform to be successfully implemented. At worst, it diverts attention from the central purpose of most attempts to take reform to scale: to improve teaching and learning for large numbers of students.

This lack of focus on the nature of change within the definition of scale itself is particularly problematic given what we know about the challenges of making change in classroom practice. The history of public schooling is replete with evidence of reforms that barely scratched the surface of schooling, failing to reach into the classroom to influence instruction (Cuban, 1988; Elmore, 1996; Elmore, Peterson, & McCarthy, 1996; Sarason, 1971; Tyack & Cuban, 1995). Furthermore, when teachers do bring reforms into their classrooms, they do so in ways that vary, at times substantially, in depth and substance (Coburn, 2002; Datnow et al., 2002; EEPA, 1990; Jennings, 1996; Spillane, 1999; Spillane & Jennings, 1997; Spillane, Reiser, & Reimer, 2002). Because teachers draw on their prior knowledge, beliefs, and experiences to interpret and enact reforms, they are likely to "gravitate" toward approaches that are congruent with their prior practices (Spillane 2000, p. 163), focus on surface manifestations (such as discrete activities, materials, or classroom organization) rather than deeper pedagogical principles (Coburn, 2002; Spillane, 2000; Spillane & Callahan, 2000; Spillane & Zeuli, 1999), and graft new approaches on top of existing practices without altering classroom norms or routines (Coburn, 2002; Cuban, 1993). All of this suggests that reaching the classroom cannot be taken for granted and that what counts as classroom implementation must be carefully defined.

To that end, I propose that the nature of change take center stage in conceptualizations of scale. More specifically, to be "at scale," reforms must effect deep and consequential change in classroom practice (see also Elmore, 1996; McLaughlin & Mitra, 2001; Olson, 1994a on this point). By "deep change," I mean change that goes beyond surface structures or procedures (such as changes in materials, classroom organization, or the addition of specific activities) to alter teachers' beliefs, norms of social interaction, and pedagogical principles as enacted in the curriculum. By *teachers' beliefs*, I am referring to teachers' underlying assumptions about how students learn, the nature of subject matter, expectations for students, or what constitutes effective instruction. Many external reform initiatives promote a view of

teaching and learning that challenges conventional beliefs about one or more of these dimensions. The question is: Do teachers' encounters with reform cause them to rethink and reconstruct their beliefs? Or do they alter reforms in ways that reinforce or reify pre-existing assumptions?

By *norms of social interaction*, I am referring to teacher and student roles in the classroom, patterns of teacher and student talk, and the manner in which teachers and students treat one another. This dimension is an important component of depth because many external reform efforts explicitly target these central aspects of classroom life. For example, reforms rooted in constructivist learning theory seek to alter traditional student and teacher roles (e.g., ATLAS and Fostering a Community of Learners). Reforms that emphasize collaborative learning seek to increase student-student talk around matters of teaching and learning (e.g., Success for All). And reforms that address the environment for learning in schools and classrooms seek to shape the nature and quality of teacher-student and student-student relationships (e.g., CDP, CES, and Comer Schools). But beyond a reform's specific focus, norms of social interaction reveal much about teachers' views about where expertise and knowledge is located in the classroom and how knowledge is developed (Bransford, Brown, & Cocking, 2000; Spillane & Zeuli, 1999; Stokes et al., 1997).[5]

Finally, and perhaps most importantly, depth involves changes in *underlying pedagogical principles* embodied in the enacted curriculum. Following Cohen and Ball (1999), I define enacted curriculum as the ways that students and teachers engage with particular materials or activities over time. Abundant research has illustrated the way pre-specified tasks and curriculum are often transformed as they are enacted (Coburn, 2001; Hoffman, et al., 1998; Spillane & Jennings, 1997; Spillane & Zeuli, 1999; Stein, Grover, & Henningsen, 1996), with implications for the nature of pedagogical approaches, representations of subject matter, and learning opportunities for students. Thus, it is important to look beyond the presence or absence of specific materials or tasks to the underlying pedagogical principles embodied in the ways teachers engage students in using these materials and tasks.

Some may argue that these components of depth are more appropriate for principle-based reforms than for materials or activity-based reforms that might have greater elaboration or even scripting. However, most if not all reforms, even those that are *not* explicitly principle based, "carry" sets of ideas about what constitutes appropriate instruction. That is, ideas about the nature of the subject matter, valued student outcomes, how students learn, and appropriate and effective pedagogy are embedded in the materials or activities. Thus, these approaches *do* put forth particular pedagogical principles, albeit in an often-implicit manner. Thus, it is appropriate to track the enactment of these pedagogical principles. In addition, many activity- or materials-based reforms, including those that are scripted, construct activities in ways that seek to alter norms of teacher-student and student-student interaction. Critics might also argue that the components of depth I propose are more appropriate for reforms that promote "ambitious pedagogy" than those promoting more conventional approaches to teaching and learning. Although I agree with theorists who suggest that it is more challenging to make consequential change in practice the further proposed practice is from existing practice (Cohen & Ball, 2000), I argue that any reform that seeks

to promote change in classroom practice can be implemented at varying degrees of depth.[6]

A conception of scale that fully incorporates depth has key implications for research design, suggesting not only an increased emphasis on measures of classroom change, but also a focus on measures that capture beliefs, norms, and pedagogical principles as enacted in the classroom. While many studies of scale do include some measure of the degree of implementation in classrooms, researchers often measure implementation in terms of the presence or absence of materials or prescribed activity structures (e.g., Berends, 2000; Berends, Kirby, Naftel, & McKelvey, 2001; Cooper et al., 1997). Other studies provide such limited methodological information about classroom implementation that it is difficult to ascertain the degree to which researchers rely upon surface-level changes in materials or activities or changes in underlying norms, beliefs, and principles (e.g., Smith et al., 1998; Supovitz, Poglinco, & Snyder, 2001). Yet, as Spillane and Jennings (1997) demonstrate, it is possible to come to very different conclusions about the degree of implementation of reform practice depending upon whether a researcher focuses on activity structures and material use versus what Spillane and Jennings call "'below-the-surface' differences in pedagogy" (p. 453). Thus, measuring deep and consequential change in classroom practice requires explicit attention to beliefs, norms, and pedagogical principles.

Existing research in reform implementation and a few studies of scale suggest a range of possible strategies for capturing these dimensions of classroom change. Although not a study of scale, researchers with the QUASAR project (Quantitative Understanding Amplifying Student Achievement and Reasoning) investigated changes in mathematics instruction over a 5-year period videotaping classroom instruction and analyzing mathematical tasks as they unfolded in teachers' classrooms. In other words, they studied the enacted curriculum, analyzing the nature of the cognitive demands and such features of the tasks as the potential for multiple solutions and the requirements for student explanation or justification (Stein, Grover, & Henningsen, 1996; Stein, Henningsen, & Grover, forthcoming). In a study of scale-up of external school reforms in 13 schools, Datnow and Yonezawa (in press) used an observation instrument that combined ethnographic field notes with a structured observation protocol. By combining these approaches, not only were they able to capture the presence or absence of design elements, such as the use of cooperative learning or alternative assessment, but they also were able to analyze the character and quality of social interaction and the enactment of pedagogical principles underlying the designs. For example, the instrument required observers to analyze ethnographic field notes to determine if and how pedagogical principles in the CES, such as "tone of decency," "teacher as coach," and "personalization," were enacted in the classroom.

The increased emphasis on depth as a key element of scale calls into question the degree to which classroom implementation can be assessed using survey methods alone. Capturing depth may require in-depth interviewing and classroom observation, refocused on such indicators as the nature of instructional tasks, discourse patterns in the classroom, and teachers' conceptions of knowledge and learning. Other methods less often used in studies of scale, such as the systematic collection of student work samples or the use of teacher logs (see, for example, Ball, Camburn,

Correnti, Phelps, & Wallace, 1999), may also have the potential to capture fundamental changes in classroom instruction.

Sustainability

As a second element of scale, consequential change must be sustained. The concept of scale primarily has meaning over time. The distribution and adoption of an innovation are only significant if its use can be sustained in original and even subsequent schools. Reforms can be adopted without being implemented, and can be implemented superficially only to fall into disuse. Yet, while the idea of sustainability is fundamental to scale-up, few conceptualizations address it explicitly. It only rarely appears in theoretical and empirical pieces (Elmore, 1996; McLaughlin & Mitra, 2001, are exceptions). Most discussions address issues of sustainability and scale separately, obscuring the way that scale, in fact, depends upon sustainability. And, perhaps most seriously, only a minority of studies of scale have employed designs that have allowed these studies to investigate sustainability. For example, of the 44 publications on efforts to scale up external reforms reviewed for this article, only 18 publications involved investigations of schools that had been involved in the reform for 4 or more years. And only one explicitly looked at schools involved in reforms for which an implementation period with additional resources and attention had officially ended. Instead, most studies focus on schools in their first few years implementing a new external reform, failing, in our view, to capture sustainability.

Yet there is ample evidence that sustainability may be the central challenge of bringing reforms to scale. Schools that successfully implement reforms find it difficult to sustain them in the face of competing priorities, changing demands, and teacher and administrator turnover (Berends, Bodilly, & Kirby, 2002; Bodilly, 1998; Comer, Haynes, Joyner, & Ben Avie, 1996; Cuban, 1993; Fink, 2000; Hargreaves & Fink, 2000; Hatch, 2000; Healy & DeStefano, 1997; Louis & Miles, 1990; MacIver, Stringfield, & McHugh, 2000; McLaughlin & Mitra, 2001; Muncey & McQuillan, 1996; Stringfield & Datnow, 1998; Tyack & Cuban, 1995). Externally developed school reforms may be especially vulnerable to this problem because implementation typically involves a short-term influx of resources, professional development, and other forms of assistance to facilitate implementation that dissipates over time as external developers turn their attention to other sites (Datnow et al., 2002; MacIver et al.; McLaughlin & Mitra; Stokes, 1997).

This suggests the need for a renewed and vigorous dialogue, not just about the challenges of sustainability, but about strategies for providing schools with the tools they will need to sustain the reform, especially after initial influx of resources dissipates. Interestingly, recent research suggests that depth may play an important role in schools' and districts' capacity to sustain change. At the classroom level, teachers with a deep understanding of the pedagogical principles of a reform are better able to respond to new demands and changing contexts in ways that are consistent with underlying principles of reform, thus sustaining and, at times, deepening reform over time (Coburn & Meyer, 1998; McLaughlin & Mitra, 2001). But because classrooms are situated in and inextricably linked to the broader school and system, teachers are better able to sustain change when there are mechanisms in place at multiple levels of the system to support their efforts. This includes the presence of a supportive professional community of colleagues in the school that reinforces normative changes and provides continuing opportunities to learn (McLaughlin & Mitra; Stokes et al., 1997), knowledgeable and supportive school leadership (Berends et al., 2002; Comer et al., 1996; Datnow et al., 2002; Fullan & Steigelbauer, 1990; Hargreaves & Fink, 2000; Legters et al., 2002; McLaughlin & Mitra; Muncey & McQuillan, 1996; Murphy & Datnow, 2003), connections with other schools or teachers engaged in similar reform (Cooper, Slavin, & Madden 1998; McDonald et al., 1999; Muncey & McQuillan), and normative coherence or alignment between the district policy context and the reform (Berends et al., 2002; Comer et al.; Datnow et al., 2002; McLaughlin & Mitra). Reformers, then, need to ask two questions: (a) Which strategies are effective at developing and nurturing depth in teachers' enactment of the reform? and (b) How can reformers work to create the key conditions in schools and districts that support and sustain classroom change over time?

The explicit focus on sustainability as a key element of scale also has implications for research design. At a most basic level, it highlights the need for designs that actually allow researchers to assess whether or not changes in schools and classrooms persist over time. Existing studies suggest a range of possible strategies to capture sustainability. In our research on the CDP, we studied schools and districts for 2 years after formal funding for the 4-year initiative had ended. Thus, the schools in the study been involved with the reform effort for 6 years. Furthermore, they had experienced a transition from a 4-year period of intensive involvement with reform developers, additional influx of resources, and some protection from the district, to one where they had to continue the reform with few additional resources or special treatment from the district. This research design allowed us to investigate whether or not and how schools sustained the reform after an external partner and funding dissipated. Other studies in the literature on scale employ designs that sample schools with a range of years of experience participating in the reform (Cooper et al., 1998; Datnow et al., 2000). For example, in a mixed-method study of implementation and sustainability of Success for All, Cooper and his colleagues (1998) sampled schools that had participated in Success for All from 1 to 8 years. In another example, in their 3-year study of the Core Knowledge program, Datnow et al. (2000) sampled schools that were new to the program and schools that had been involved in the program for at least 2 years at the start of the study that were identified as "advanced" in their implementation. At a minimum, these designs allow researchers to be certain that studies of scale actually included schools that moved beyond initial implementation to sustain external reform efforts over time. But these designs could also potentially provide researchers a rich opportunity to explore if and how the challenges and processes involved with sustaining reform efforts differ at different stages of a school's experience with the reform. We know a lot about challenges to sustainability in the early years of reform. But how do these challenges differ as reforms mature and initial energy, personnel, and funding dissipate? Research designs such as these become vitally important as issues of sustainability move to the forefront of conceptualizations of scale.

Spread

As should be clear, the central insight of traditional definitions of scale—the spreading of reform to greater numbers of classrooms and schools—remains the core of the expanded conceptualization offered here. I have already discussed the importance of taking into consideration *what* is spread, suggesting that scaling up must involve more than the spread of activity structures, materials, and classroom organization; it must also involve the spread of underlying beliefs, norms, and principles to additional classrooms and schools. But this more explicit focus on norms and principles also suggests a modification in the notion of spread itself. Rather than thinking of spread solely in terms of expanding outward to more and more schools and classrooms, this emphasis on the normative highlights the potential to spread reform-related norms and pedagogical principles *within* a classroom, school, and district. For example, at the district level, spread not only involves increasing the number of schools that participate, but also the ways in which reform norms and principles influence district policies, procedures, and professional development (see, also, Comer et al., 1996, on this point). Spread at the school level not only involves the reform moving to more and more classrooms, but also reform principles or norms of social interaction becoming embedded in school policy and routines. At the classroom level, a reform can spread within as teachers begin to draw on reform norms and principles in aspects of their practice beyond specific reform-related activities or subject matter (McLaughlin & Mitra, 2001).

While I do not mean to suggest that all reformers do or should seek to foster spread within, recognizing this aspect of spread may be especially important given evidence of the key role of normative coherence at the school, in district, and in the broader environment in sustaining reform. As discussed earlier, teachers and schools are more likely to be able to sustain and deepen reform over time when school and district policy and priorities are compatible or aligned with reform. Spreading reform norms, beliefs, and principles within schools and districts may be a key mechanism for developing this normative coherence (Coburn & Meyer, 1998). Thus, this type of spread may be especially important for reforms that challenge conventional or institutionalized approaches to teaching and learning in significant ways.

Broadening notions of spread to include spread within the system raises strategically different issues for reformers. For example, it suggests a different way to think about the role of the district in spreading a reform effort. Typically, reformers have focused on the degree to which the district can provide support or protection for schools implementing a reform (Bodilly, 1998; Hatch, 1998; McDonald et al., 1999; Olson, 1994b; Slavin & Madden, 1996). Recasting spread to include spread of norms and principles within suggests that the district's role may be important beyond the support it provides to schools or as a way to create geographic proximity; the district may be a strategic site for spread itself. External reformers might then focus on threading reform ideas throughout the district office, creating knowledgeable leaders who can influence policy, procedures, and values.

Taking into account "spread within" also has implications for research. If we broaden notions of spread to include "spread within," we must also broaden the indicators used to measure spread. However, because research on scale has so rarely explored this dimension of spread, there are few models to draw upon for guidance. One strategy is to draw on wisdom from implementation research that investigates the enactment of pedagogical principles and shifting norms of interaction at the classroom level and apply this wisdom to different levels of the system. For example, to capture spread within the district, researchers could borrow the notion of enacted curriculum and track the degree to which pedagogical principles of the reform are embedded in district policy and enacted in district professional development. Corcoran's (2003) study of the Merck Institute for Science Education's (MISE) work in four districts provides an illustration of this approach. He documents how collaboration with MISE influenced science frameworks, teacher observation procedures, allocation of instructional resources, and approaches to professional development in participating districts. Researchers could also investigate shifts in district's standard operating procedures over time (e.g., procedures for curriculum decisions, ongoing interaction with school leaders, roles for teacher leaders) to ascertain the degree to which the reforms influence these norms of interaction.[7] Similar strategies might also be applicable at the school level.

Finally, to capture "spread within" at the classroom level, researchers could investigate the degree to which teachers draw on pedagogical principles and norms of interaction in areas of the classroom beyond those subjects, times of day, or particular activities targeted by reform (see McLaughlin & Mitra, 2001, for an example of this strategy). In broadening indicators of spread in this way, researchers have the potential to capture aspects of spread that may prove important for understanding the influence of external reforms on schools and systems.

Shift in Reform Ownership

Finally, to be considered "at scale," ownership over the reform must shift so that it is no longer an "external" reform, controlled by a reformer, but rather becomes an "internal" reform with authority for the reform held by districts, schools, and teachers who have the capacity to sustain, spread, and deepen reform principles themselves. In most cases, when schools embark on external reform efforts, the knowledge and authority for the reform is situated outside the school, usually with the outside provider responsible for spreading the reform. Yet, ultimately, it is the teacher's, school's, and, in some cases, district's responsibility to enact and sustain the reform in ways that make a difference to students. One of the key components of taking a reform to scale, then, is creating conditions to shift authority and knowledge of the reform from external actors to teachers, schools, and districts. Or, as Stokes et al. (1997) argue, the reform must transition from "an externally understood and supported theory to an internally understood and supported theory-based practice" (p. 21). With this shift in ownership, the reform becomes self-generative (McLaughlin & Mitra, 2001).

Most conceptualizations of scale are virtually silent on the shift from external to internal (McLaughlin & Mitra, 2001, are the exception). Reformers and researchers who talk about the importance of "reform ownership" (see, for example, Cooper, 1998; Datnow & Castellano, 2000; Legters et al., 2002; Olson, 1994b; Slavin, 1997; Slavin & Fashola, 1998) often understand this concept as "buy-in" or acceptance, rather than a shift in knowledge

of and authority for the reform. These discussions of ownership are also much more likely to concern initial adoption and implementation (e.g., how to ensure teacher "buy-in") rather than long-term sustainability and growth (e.g., how to develop capacity for school faculty to take the initiative to maintain reform in the face of constantly changing priorities). In addition, reformers have tended to have been much more concerned about how to develop their own capacity to continue to provide professional development and technical assistance as reforms spread (Bodilly, 1998; Olson, 1994b; Slavin & Madden, 1996, 1999; Stringfield & Datnow, 1998)—a task that grows geometrically with greater and greater spread—than how to work with schools and districts to develop the capacity necessary for them to assume authority and knowledge for the reform.

Although this component of scale is rarely explicitly incorporated into studies of attempts to take external reform efforts to scale, there are a few studies that provide hints about what such a transition might entail. Davila and Gomez (1995) describe a scale-up strategy that hinged on developing a cadre of knowledgeable teacher leaders who, over time, assumed responsibility for providing ongoing professional development to teachers new to the reform (see, also, Coburn & Meyer, 1998, on this point). Given high rates of teacher and administrator turnover in some communities as well as experienced teachers' needs to deepen their approach over time (Muncey & McQuillan, 1996), it is likely that teachers and administrators will need ongoing professional development to sustain, deepen, and spread reform. This suggests that developing the capacity to provide reform-related professional development or other structures for ongoing teacher and administrator learning may be a central feature of shifting authority and ownership for the reform.

In another study, McLaughlin and Mitra (2001) point out that shift in reform ownership also requires transferring substantive and strategic decision making from the reform organization to district and school leaders. This shift requires that reformers cultivate deep, reform-centered knowledge among key leadership and model ways to draw upon that knowledge in ongoing decision making (McLaughlin & Mitra; Stokes et al., 1997). Teachers, school leadership, and district leadership need to exercise this reform-centered decision making as they work to sustain practice in the face of new circumstances, initiatives, and priorities that may or may not conflict with reform. For example, our study of the CDP suggests that leaders with in-depth understanding of reform principles were better able to interrogate new policy and reform initiatives to ascertain the degree to which they were coherent with the CDP, rejecting or altering initiatives that they deemed incongruent and weaving the CDP into new initiatives and priorities that appeared congruent (Coburn & Meyer, 1998). Furthermore, schools and districts may need this depth in substantive and strategic decision making if they are to fully take on the responsibility of spreading reform over time.

Finally, shift in ownership may require that schools and districts develop the capacity to generate continued funding for reforms. Many reform efforts are supported by external grant funds or policy initiatives that are finite. Yet, activities such as ongoing professional development and other efforts to spread and deepen reform also require funding. Thus, reform ownership may re-

quire the capacity to think creatively about reallocating existing funding streams and/or the capacity to locate and secure additional grant funding to support activities that deepen and spread reform over time (Coburn & Meyer, 1998).

All of this suggests that depth of reform-centered knowledge—not only at the classroom level but also among leaders at multiple levels of the system—is a key element in shifting ownership and authority of reform. It further suggests that shift in reform ownership may be a central element in sustaining and spreading reform in the face of shifting priorities, changes in funding, and challenges to policy coherence.

Placing reform ownership as a central element of scale raises the priority for directing reform attention and resources to strategies that have the potential for enabling schools and districts to assume ownership for the reform over time. To date, discussions of the shift from external to internal have been relatively absent in the literature, and yet, there are many strategic questions to consider. What strategies are effective in cultivating the capacities necessary to assume authority for reform? Are these strategies different at different levels of the system (classroom, school, district)? Should they vary for different kinds of reforms? How can reformers lay the groundwork for a shift from external to internal from the early days of engagement with a school or district?

This reconceptualization of scale also has implications for researchers. Shift in reform ownership has rarely been incorporated into studies of attempts to take external reform efforts to scale; thus, it represents a new outcome for studies of scale. Existing research suggests several preliminary indicators for shift of reform ownership at the school and district levels: (a) the presence of structures and mechanisms for ongoing teacher learning about reform (e.g., professional development, teacher study groups); (b) the presence of established strategies to provide continued funding for reform activities; (c) the degree to which districts have taken responsibility for continued spread of reform; and (d) the use of reform-centered ideas or structures in school or district decision making. However, more research is clearly needed to elaborate, extend, and validate these indicators.

Discussion

The problem of scale remains one of the most pressing issues in educational reform and improvement. In an effort to capture the multidimensional nature of the problem, I offer an elaborated conceptualization of scale that requires that reform not only reach more widely but also more deeply into schools to effect and sustain consequential change. It emphasizes the spread of norms, beliefs, and pedagogical principles both between and within classrooms, schools, and districts. And it includes an additional outcome—the shift in ownership—that may prove key to schools' and districts' abilities to sustain and spread the reform over time. By highlighting depth, sustainability, spread, and ownership, this expanded conceptualization brings them to the forefront of discussions of reform strategy, articulating goals and raising questions about effective approaches to help schools and districts achieve these goals. It also points to the need for new research designs better suited to capture this more complex vision.

Broadening the definition of scale in this way, however, also highlights inherent tensions for both researchers and reformers. For researchers, this conceptualization emphasizes dimensions of

scale that are more challenging to measure. It is more challenging to measure conceptual change or enacted pedagogical principles than the presence or absence of activities or materials. It is more challenging to measure the spread of norms of interaction than the number of teachers or schools involved in an initiative. And it is arguably more challenging to measure the shift in authority over and knowledge of reform than reform adoption and sustainability. There are also tradeoffs in resources, time, and effort in studying breadth versus depth. Research strategies that capture depth and shift in ownership, most often qualitative, tend to be more expensive and time consuming than survey and other quantitative methods better suited to capture breadth.

But it seems important to wrestle with these challenges to ensure that we develop research designs that capture what is important rather than only what is easily measurable.[8] To that end, we need continued conceptual and methodological development to identify and validate measures of such heretofore neglected dimensions of scale as spread within schools and districts, and shift in reform ownership at multiple levels of the system. We need to explore creative and cost-effective ways to study schools that have been engaged in reform initiatives for more than a few years. And if our argument is persuasive that we need to investigate the multiple dimensions of scale to fully understand the long-term dynamics and success of reform scale-up, we need a continued dialogue on how to strike the appropriate balance between depth and breadth.

For reformers, there are also tensions. Like researchers, reform organizations must navigate the tension between breadth and depth. The capacity building at multiple levels of the system that may be necessary for depth and reform ownership is likely to be expensive and resource-intensive, which may limit developers' ability to expand as broadly (see Comer et al., 1996; McDonald et al., 1999; Muncey & McQuillan, 1996; Slavin & Madden, 1999 for further discussion on this point). This tension may grow increasingly acute the more the reform diverges from existing practice and the more complex or comprehensive the reform's approach and goals (Cohen & Ball, 2000; Cuban, 1988). That is, the more challenging a reform is to teachers' existing beliefs and practices, or the more aspects of classroom practice or levels of the system it engages, the more it may need well-elaborated materials and sustained, ongoing professional development to achieve depth.[9] Similarly, reforms of this nature may require more effort on the part of reformers to work with multiple levels of the system to encourage normative coherence and sustainability.[10] This suggests that the more ambitious a reform, the more challenging it may be to simultaneously achieve spread, sustainability, and depth.

It is possible that some of the tension between depth and breadth can be mitigated through the design of the reform itself. For example, the more that knowledge and guidance is built into the reform via greater elaboration or even scripting, the less reformers may need to invest in professional development. But this approach elicits other tensions. Absent some mechanism for teachers and others in the system to learn pedagogical principles and norms from the reform, it may be difficult for them to develop the deep knowledge and authority necessary for reform ownership. Similarly, this approach does not address the challenges of normative coherence at multiple levels of the system.

Finally, there are also tensions between reform ownership and fidelity, especially for reforms that place a high priority on fidelity to particular activity structures. As knowledge and authority shifts from external reform organizations to school and district personnel, the decisions about what aspects of the reform to emphasize or adapt no longer lie with the external reform organization. For reform organizations that advocate fidelity to underlying norms or principles, this is not troubling as long as school and district personnel have deep reform-centered knowledge. With such knowledge, teachers and others will theoretically be able to make decisions about the reform in ways that remain faithful to the underlying philosophy and pedagogical principles, thus mitigating some of the tension between reform ownership and fidelity (Brown & Campione, 1996; McLaughlin & Mitra, 2001). If, however, precise activity structures are key to the reform, as is sometimes the case with many reforms that advocate a replication approach to spread, the tension may prove difficult to navigate.

The research agenda on scale is still in the process of formation. With the growing number of external reforms with a longer history of development and use, conditions are ripe for studies that grapple with the challenges of creating research design to further explore the multidimensionality of scale. With attention to this multidimensionality, research can begin to speak more clearly and persuasively about the tensions and tradeoffs involved in different strategies to take reform to scale.

NOTES

[1] Bodily (1998) and Glennan (1998) provide a good illustration of this phenomenon with the New American Schools (NAS). NAS conceived of scale as increasing the numbers of schools in a bounded geographic area to create a critical mass. This construction influenced their decision to work closely with what they call "jurisdictional operating environments," most often districts, to spread reform models. In another vein, it is likely that reformers who view scale as replication of reform (e.g., Robert Slavin and Success for All) are apt to make strategically different choices in the nature of the materials, professional development, and strategies for follow-up than those reformers who conceptualize scale in terms of co-creation or mutual adaptation (e.g., Ted Sizer and the Coalition of Essential Schools) (McDonald et al., 1999; Olson, 1994b; Slavin & Madden, 1994, 1996).

[2] To identify literature on scale, I searched the ERIC database (1992–2002) using the following descriptors: "scaling up," "scale-up," "scale" and "reform," and "scale" and "innovation." (A search simply using the term "scale" was not practical for the ERIC database as it elicited 7,097 hits, many of which had to do with measurement issues.) I also searched the National Clearinghouse for Comprehensive School Reform (www.goodschools.gwu.edu/) using the descriptors "scale," "scaling up," and "scale-up." I then reviewed reference lists from the resulting studies, yielding many additional citations. Finally, to address reviewers' concerns about the limitations of these databases in identifying books, I searched the publication lists of the top 10 scholarly publishers

who publish books on educational research and reform. From all these searches, I selected only those pieces to review that met the following criteria. First, publications needed to be focused on the processes or outcomes of attempts to scale up external school reform efforts focused on improving classroom instruction. Thus, I excluded articles focused on state and district policy (such as standards and accountability programs) and the scale up of nonclassroom interventions (such as one-on-one tutoring programs, governance structures, or school-community collaborations). Second, because I was interested in several different genres, I selected pieces that were empirical studies, position papers or conceptual pieces, and descriptive accounts of reform strategies. In all, 44 books, articles, or book chapters from 19 different empirical studies, 18 position papers or conceptual studies, and 10 descriptive accounts of reform strategy met these criteria and, therefore, were included in the review.

[3] The Child Development Project (CDP), a program of the Developmental Studies Center (DSC) in California, is a whole school reform program for elementary schools that emphasizes developing the social, ethical, and intellectual dimensions of learning among children (Battistich, Schaps, Watson, & Solomon, 1996; Battistich, Solomon, & Watson, 1998; Developmental Studies Center, no date; Watson, Battistich, & Solomon, 1997). In 1991, after 10 years of developing their model through close collaboration with two schools, the DSC engaged in a 4-year effort to bring the CDP to 12 schools in six districts. We began a study of the CDP in 1995, the year after the end of formal funding and the formal relationship between the DSC and the districts ended. Using a mixed-methods design, we followed a subset of schools and districts for 2 years after the end of formal funding. We then linked our data with data collected by DSC researchers during the 4 years of the project to track teachers', school leaders', and district personnel's experiences with the reform over 6 years (see Coburn & Meyer, 1998; and Stokes et al., 1997, for additional information on the study).

[4] I limit the discussion to scaling up to external school- or classroom-based reforms focused on classroom instruction as most of the evidence on scale-up is rooted in studies of external reform initiatives. It is possible that the conceptualization I offer here may have relevance to issues of scale with policy initiatives. But given substantive differences in mechanisms of spread, funding, and authority relations with policy, it seems premature to extend the conceptualization beyond external reform initiatives absent additional evidence of its applicability.

[5] It is important to note that norms of social interaction may be influenced by classroom organization (ways of grouping children) or activity structures (e.g., guided reading groups or writers' workshops). However, altering these elements does not necessarily result in shifting norms of interaction. For example, cooperative learning groups have the intention of increasing student-to-student interaction around matters of teaching and learning. Yet, as many researchers have pointed out, just because students are seated in cooperative groups does not mean that they are actually working cooperatively (Datnow et al., 2002; Datnow & Yonezawa, in press; Spillane & Jennings, 1997).

[6] In this way, our conception of depth diverges from Hargreaves and Fink (2000) who, in defining depth as learning for social and emotional understanding, put forth a conception that is more rooted in constructivist or "ambitious" pedagogy.

[7] Of course, the specific indicators that one would track would depend upon the focus of reform. For example, with Comer's School Development Project, which advocates for the use of collaborative and inclusive decision making as a central aspect of the reform (Comer et al., 1996; Cook, Habib, Phillips, Settersen, Shagle, & Degirmenciougku, 1999; McDonald et al., 1999), researchers could conceivably track changes in norms of interaction in district decision-making processes to evaluate the degree to which they reflect project norms. In another example, the CDP embraced a constructivist and situated approach to teacher professional development (Dasho & Kendzior, 1995; Develop-

mental Studies Center, no date; Lewis, Watson, & Schaps, 1997). In our research on the CDP, we investigated the degree to which districts drew on these pedagogical principles as they designed their own opportunities for continued professional development for teachers and principals. This, then, served as a key indicator of spread within the district for this reform.

[8] This point was inspired by recent writing (e.g., Berends et al., 2002) on the challenges of accurately measuring student learning and achievement.

[9] See Cohen and Ball 2000 for a more extended discussion of this point.

[10] Some reformers have also argued that concentrating spread within a district is one way to reduce the human and fiscal resources devoted to managing the district policy environment (Comer et al., 1996) while increasing the likelihood of sustainability and spread (see Bodilly, 1998).

REFERENCES

Ball, D. L, Camburn, E., Correnti, R., Phelps, G., & Wallace, R. (1999). *New tools for research on instruction: A web-based teacher log.* Working paper, Center for Teaching Policy. Seattle: University of Washington.

Battistich, V., Schaps, E., Watson, M., & Solomon, D. (1996). Prevention Effects of the Child Development Project: Early Findings from an Ongoing Multisite Demonstration Trial. *Journal of Adolescent Research, 11,* 12–35.

Battistich, V., Solomon, D., & Watson, M. (1998, April). *Sense of community as a mediating factor in promoting children's social and ethical development.* Paper presented at the annual conference of the American Educational Research Association, San Diego, CA.

Berends, M. (2000). Teacher-reported effects of New American Schools designs: Exploring relationships to teacher background and school context. *Educational Evaluation and Policy Analysis*, 22(1), 65–82.

Berends, M., Bodilly, S., & Kirby, S. N. (2002). *Facing the challenges of whole-school reform: New American Schools after a decade.* Santa Monica, CA: RAND.

Berends, M., Kirby, S. N., Naftel, S., & McKelvey, C. (2001). *Implementation and performance in New American Schools: Three years into scale-up.* Santa Monica, CA: RAND.

Blum, R. E. (1997, March). *Barriers to scaling up school change.* Paper presented at annual conference of the American Education Research Conference, Chicago, IL.

Bodilly, S. (1998). *Lessons from New American Schools' scale-up phase: Prospects for bringing designs to multiple sites.* Santa Monica, CA: RAND.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Research Council.

Brown, A., & Campione, J. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. 289–325). Mahwah, NJ: Erlbaum.

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. (2002, August). *Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom.* Paper presented at the American Sociological Association Conference, Chicago, IL.

Coburn, C. E., & Meyer, E. R. (1998, April). *Shaping context to support and sustain reform.* Paper presented at the American Educational Research Association Conference, San Diego, CA.

Cohen, D. K., & Ball, D. L. (1999). *Instruction, capacity, and improvement.* CPRE Research Report Series RR-43. Philadelphia: Consortium for Policy Research in Education.

Cohen, D. K., & Ball, D. L. (2000). *Instructional innovation: Reconsidering the story*. Working paper for the Study of Instructional Improvement. Ann Arbor: University of Michigan.

Comer, J., Haynes, N., Joyner, E. T., & Ben Avie, M. (1996). *Rallying the whole village: The Comer process for reforming education*. New York: Teachers College Press.

Cook, T. D., Habib, F., Phillips, M., Settersten, R. A., Shagle, S. C., & Degirmenciouglu, S. M. (1999). Comer's School Development Program in Prince George's County, Maryland: A theory-based evaluation. *American Educational Research Journal, 36*(3), 543–597.

Cooper, R. (1998). *Socio-cultural and within-school factors that affect the quality of implementation of schoolwide programs* (Report No. 28). Baltimore: Center for Research on the Education of Students Placed At Risk, Johns Hopkins University.

Cooper, R., Slavin, R. E., & Madden, N. (1997). *Success for All: Exploring the technical, normative, and socio-cultural dimensions of scaling up* (Report Number 16). Baltimore: Center for Research on the Education of Students Placed At Risk, Johns Hopkins University.

Cooper, R., Slavin, R. E., & Madden, N. (1998). Success for All: Improving the quality of implementation of whole-school change through the use of a national reform network. *Education and Urban Society, 30*(3), 385–408.

Corcoran, T. (2003). *The Merck Institute for Science Education: A successful intermediary for education reform*. Philadelphia: Consortium for Policy Research in Education.

Cuban, L. (1988). Constancy and change in schools (1880s to the present). In P. Jackson (Ed.), *Contributing to educational change* (pp. 85–105). Berkeley, CA: McCutchan.

Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms 1890–1990* (2nd ed.). New York: Teachers College Press.

Dasho, S., & Kendzior, S. (1995, April). *Toward a caring community of learning for teachers: Staff development to support the Child Development Project*. Paper presented at the annual conference of the American Educational Research Association, San Francisco, CA.

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' response to Success for All: How beliefs, experiences, and adaptations shape implementation. *American Educational Research Journal, 37*(3), 775–799.

Datnow, A., Hubbard, L., & Mehan, H. (2002). *Extending educational reform: From one school to many*. London: Routledge/Falmer.

Datnow, A., Stringfield, S., McHugh, B., & Hacker, B. (1998). Scaling up the Core Knowledge Sequence. *Education and Urban Society, 30*(3), 409–432.

Datnow, A., & Yonezawa, S. (in press). Observing school restructuring in multilingual, multiethnic classrooms: Balancing ethnographic and evaluative approaches. In H. Waxman, R. Tharp, & R. Hilberg (Eds.), *Observational research in culturally and linguistically diverse classrooms*. Cambridge, UK: Cambridge University Press.

Davila, N., & Gomez, M. (1995, April). *Evaluation of school-based regional dissemination centers as scale-up mechanism for systemic educational reform in science and mathematics*. Paper presented at the annual conference of the American Educational Research Association, San Francisco, CA.

Developmental Studies Center. (no date). *A vision for American education: Schools as caring communities of learners. DSC's strategic plan*. Developmental Studies Center: Unpublished Manuscript.

EEPA (1990). *Educational evaluation and policy analysis. Entire volume. 12*(1), 233–353.

Elmore, R. F. (1996). Getting to scale with successful educational practices. *Harvard Educational Review, 66*(1), 1–26.

Elmore, R. F., Peterson, P. L., & McCarthey, S. J. (1996). *Restructuring in the classroom: Teaching, learning, and school organization*. San Francisco: Jossey-Bass.

Fink, D. (2000). *Good schools/real school: Why school reform doesn't last*. New York: Teachers College Press.

Fuchs, D., & Fuchs, L. S. (1998). Researchers and teachers working together to adapt instruction for diverse learners. *Learning Disabilities Research & Practice, 13*(3), 126–137.

Fullan, M. (2000). The return of large-scale reform. *Journal of Educational Change, 1*, 1–23.

Fullan, M., & Stiegelbauer, S. (1991). *The new meaning of educational change*. New York: Teachers College Press.

Gamson, D. (1998). *The challenge of going to scale: A review of the literature*. Paper presented at the Pew Forum, Pittsburgh, PA.

Glennan, T. K. (1998). *New American Schools after six years*. Santa Monica, CA: Rand.

Hargreaves, A., & Fink, D. (2000). The three dimensions of reform. *Educational Leadership, 57*(7), 30–34.

Hatch, T. (1998). The differences in theory that matter in the practice of school improvement. *American Educational Research Journal, 35*(1), 3–31.

Hatch, T. (2000). What does it take to break the mold? Rhetoric and reality in New American Schools. *Teachers College Record, 102*(3), 561–589.

Healy, F. H., & DeStefano, J. (1997). *Education reform support: A framework for scaling up school reform*. Research Triangle Park, NC: Research Triangle Institute.

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.

Honig, B. (1994). How can Horace best be helped? *Phi Delta Kappan, 75*(10), 790–796.

Hubbard, L., & Mehan, H. (1999). Scaling up an untracking program: A co-constructed process. *Journal of Education for Students Placed at Risk, 4*(10), 83–100.

Jennings, N. E. (1996). *Interpreting policy in real classrooms: Case studies of state reform and teacher practice*. New York: Teachers College Press.

Klein, S. P., McArthur, D. J., & Stecher, B. M. (1995, February). What are the challenges to 'scaling up' reform? In *Joining forces: Spreading strategies*. Proceedings of the Invitational Conference on Systemic Reform. Washington, DC.

Legters, N. E., Balfanz, R., Jordan, W. J., & McPartland, J. M. (2002). *Comprehensive reform for urban high schools: A talent development approach*. New York: Teachers College Press.

Lewis, C. C., Watson, M. S., & Schaps, E. (1997, March). *Conditions for school change: Perspectives from the Child Development Project*. Paper presented at the annual conference of the American Education Research Association, Chicago, IL.

Louis, K. S., & Miles, M. (1990). *Improving the urban high school: What works and why*. New York: Teachers College Press.

MacIver, M. A., Stringfield, S., & McHugh, B. (2000). *Core Knowledge Curriculum: Five-year analysis of implementation and effects in five Maryland schools* (Report No. 50). Baltimore: Center for Research on the Education of Students Placed At Risk, Johns Hopkins University.

McDermott, K. A. (2000). Barriers to large-scale success of models for urban school reform. *Educational Evaluation and Policy Analysis, 22*(1), 83–93.

McDonald, J. P., Hatch, T., Kirby, E., Ames, N., Haynes, N. M., & Joyner, E. T. (1999). *School reform behind the scenes: How ATLAS is shaping the future of education*. New York: Teachers College Press.

McLaughlin, M. W., & Mitra, D. (2001). Theory-based change and change-based theory: Going deeper and going broader. *Journal of Educational Change*, *2*(4), 301–323.

Mead, J. V., & Simon, S. (1996, April). *Developing a mid-range theory to make sense of scaling-up school reform: A cross-site case study of Chicago public schools*. Paper presented at the annual conference of the American Educational Research Association, New York City.

Muncey, D. E., & McQuillan, P. J. (1996). *Reform and resistance in schools and classrooms: An ethnographic view of the Coalition of Essential Schools*. New Haven, CT: Yale University Press.

Murphy, J., & Datnow, A. (2003). *Leadership lessons from comprehensive school reforms*. Thousand Oaks, CA: Corwin Press.

Olson, L. (1994a, November 2). Growing pains [On Assignment]. *Education Week*, pp. 31–41.

Olson, L. (1994b, November 2). Learning their lessons: Some of America's leading school reform experts reflect on what they've learned about 'going to scale.' [On Assignment] *Education Week*, pp. 43–46.

Sarason, S. (1971). *The culture of school and the problem of change*. Boston: Allyn and Bacon.

Slavin, R. E. (1997). *Sands, bricks, and seeds: School change strategies and readiness for reform*. Baltimore: Center for Research on the Education of Students Placed At Risk, Johns Hopkins University.

Slavin, R. E., & Fashola, O. S. (1998). *Show me the evidence! Proven and promising programs for America's schools*. Thousand Oaks, CA: Corwin Press.

Slavin, R. E., & Madden, N. A. (1994). Whenever and wherever we choose: The replication of Success for All. *Phi Delta Kappan*, *75*, 639–647.

Slavin, R. E., & Madden, N. A. (1996). *Scaling up: Lessons learned in the dissemination of Success for All* (Report No. 6). Baltimore: Center for Research on the Education of Students Placed At Risk, Johns Hopkins University.

Slavin, R. E., & Madden, N. A. (1999). *Disseminating Success for All: Lessons for policy and practice* (Report No. 30). Baltimore: Center for Research on the Education of Students Placed At Risk, Johns Hopkins University.

Smith, L., Ross, S., McNelis, M., Squires, M., Wasson, R., Maxwell, S. L., et al. (1998). Memphis Restructuring Initiative: Analysis of activities and outcomes that affect implementation success. *Education and Urban Society*, *30*(3), 276–325.

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). Cognition and policy implementation: District policymakers and the reform of mathematics education. *Cognition and Instruction*, *18*(2), 141–179.

Spillane, J. P., & Callahan, K. A. (2000). Implementing state standards for science education: What district policy makers make of the hoopla. *Journal of Research in Science Teaching*, *37*(5), 401–425.

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), 439–481.

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.

Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reform. *Educational Evaluation and Policy Analysis*, *21*(1), 1–27.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, *33*(2), 455–488.

Stein, M. K., Henningsen, M., & Grover, B. W. (forthcoming). Classroom instructional practices in QUASAR classrooms. In *Teaching mathematics for a change*.

Stokes, L. (1997). Short-term policy support for long-term school change: A dilemma for reform-minded practitioners. *Journal of Education Policy*, *12*(5), 371–384.

Stokes, L. M., Sato, N. E., McLaughlin, M. W., & Talbert, J. E. (1997). *Theory-based reform and problems of change: Contexts that matter for teachers' learning and community* (R97-7). Stanford, CA: Stanford University.

Stringfield, S., & Datnow, A. (1998). Scaling up school restructuring designs in urban schools. *Education and Urban Society*, *30*(3), 269–276.

Stringfield, S., Datnow, A., Ross, A., & Snively, F. (1998). Scaling up school restructuring in multicultural, multilingual contexts: Early observations from Sunland County. *Education and Urban Society*, *30*(3), 326–357.

Supovitz, J. A., Poglinco, S. M., & Snyder, B. A. (2001). *Moving mountains: Successes and challenges of the America's Choice comprehensive school reform design*. Philadelphia: CPRE.

Taylor, L., Nelson, P., & Adelman, H. (1999). Scaling-up reforms across a school district. *Reading & Writing Quarterly*, *15*, 303–325.

Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia*. Cambridge, MA: Harvard University Press.

Watson, M., Battistich, V., & Solomon, D. (1997). Enhancing students' social and ethical development in schools: An intervention program and its effects. *International Journal of Educational Research*, *27*, 571–586.

AUTHOR

CYNTHIA E. COBURN is Assistant Professor, Administrative and Policy Studies, School of Education, and Research Scientist, Learning Research and Development Center, University of Pittsburgh, 808 LRDC, 3939 O'Hara Street, Pittsburgh, PA 15260; cecoburn@pitt.edu. Her research interests include the relationships between policy and practice, school and teacher change, and qualitative research methods.