What Does It Take to Scale Up and Sustain Evidence-Based Practices?

JANETTE K. KLINGNER
ALISON G. BOARDMAN
University of Colorado Boulder
KRISTEN L. MCMASTER
University of Minnesota

ABSTRACT: This article discusses the strategic scaling up of evidence-based practices. The authors draw from the scholarly work of fellow special education researchers and from the field of learning sciences. The article defines scaling up as the process by which researchers or educators initially implement interventions on a small scale, validate them, and then implement them more widely in real-world conditions. Examples of scale-up research are included. The authors discuss challenges to scaling up and sustaining evidence-based practices, followed by factors that can potentially support scaling up, including professional development and district leadership. A case example describes how these issues can play out by highlighting experiences with a Collaborative Strategic Reading (CSR) scale-up research project in a large urban school district. The article concludes by offering recommendations for research, policy, and practice.

To effect the widespread adoption of evidence-based practices (EBPs), change is unlikely to occur one teacher—or even one school—at a time. Rather, researchers must strategically and systematically scale up implementation of EBPs in collaboration with district partners. It no longer makes sense for researchers to gather with one another to identify what they think is an important problem, write a research proposal, obtain funding to support their research, find schools and identify teachers to participate in their study, and conduct their research without substantial collaboration with the educators and leaders in their local school districts. To scale up and sustain the use of EBPs, researchers must work closely with their school district partners, not as an afterthought, but early in the process. For innovations to take hold in a district, they must meet the district’s needs and be responsive to local contextual factors. As Cobb and Smith (2008) advise, researchers should view “teachers’ instructional...
DEFINING SCALING UP

Scaling up generally refers to the process by which researchers and educators initially implement interventions on a small scale, validate them, and then implement them more widely in real-world conditions (Odom, 2009). Scaling up, as described by the Institute of Education Sciences (IES) in previous requests for applications (RFAs; most recently, 2011) focused on the process of expanding the number of schools and/or districts using an intervention and holding everything else constant while testing the effectiveness of the intervention. Perhaps because the complexities of scaling up are becoming increasingly apparent to educational researchers, the 2012 IES RFA expanded its parameters for scaling-up research to include a focus on “understanding the organizational conditions needed to support the intervention” and “determining the effects of selected moderators of the intervention” (p. 58).

In addition to evaluating the effectiveness of an intervention, researchers studying scale-up efforts commonly consider what it takes to expand and sustain an intervention in real-world settings (Fixsen & Blasé, 2009). Coburn (2003) indicated that scaling up requires four components: depth, sustainability, spread, and a shift in ownership. To be successful, scaled-up reforms must bring about deep and lasting change that goes beyond surface structures or procedures (Coburn, 2003). This transformation is consequential and sustainable over time, exists after leaders in the original schools or districts leave, and should spread to additional schools. Coburn noted that sustainability might be the most significant challenge to scaling up. Typically, after researchers depart and funding has ended, implementation wanes (Vaughn, Klingner, & Hughes, 2000). Even schools that have been able to implement reforms successfully find that sustaining them is difficult when the schools confront competing priorities, changing demands, and teacher and administrator turnover. Thus, scaling up must involve more than the spread of the surface-level aspects of a new approach, such as the routines, activities, and materials associated with it. Scaling up also requires the proliferation of the beliefs, norms, and principles underlying the approach. Finally, Coburn’s definition emphasizes that to be truly “at scale,” ownership of the practice must shift so that others no longer perceive it to be an externally driven initiative that outsiders control; but it instead becomes an internally managed effort, maintained by the districts, schools, and teachers who are implementing it. This ownership is more than tacit buy-in or acceptance but requires deeper, broader, and more substantial endorsement (McLaughlin & Mitra, 2001).

After researchers depart and funding has ended, implementation wanes.

Dunlap, Sugai, Lewis, Goodman, and Horner’s (2009) four phases of implementation when scaling up an EBP complement Coburn’s (2003) definition: (a) emergence, (b) demonstration of capacity, (c) elaboration, and (d) system adoption and sustainability. Emergence, the first phase of scaling up, takes place when the researchers or developers of a new practice decide that it might actually be scalable. The demonstration phase of scaling up is when researchers determine whether the practice is feasible and whether it has a significant effect on target outcomes. Elaboration occurs when the researchers implement the practice more broadly, drawing on lessons learned during the demonstration phase of their work and building the capacity of school district partners to implement the practice. Finally, system adoption and sustainability require the integration of the practice into the normal routines of a district or school so that the practice continues over time.

THE PROBLEM OF SCALE

The extent to which instructional innovations persist after funding has ended and researchers have departed is generally low even when innovations seem to have been effective (Cobb & Smith, 2008; Cohen & Ball, 2007; Penuel, Fishman, Cheng, & Sabelli, 2011). Many initiatives appear...
to flourish when implemented on a small scale with a few teachers or schools, but institutionalizing them so that they become part of a district's instructional culture presents numerous challenges (Fogelman, Fishman, & Krajcik, 2006). Although schools face external pressures to improve, classroom teaching seems to be remarkably stable (Cobb & Smith, 2008). This difficulty sustaining new practices may occur because researchers have not sufficiently attended to local contextual features or laid a foundation for transferring ownership of the initiative to teachers and to the district (Elmore, 1996).

The problem of scale has historically been the domain of educational policy researchers operating separately from validation researchers, innovation developers, administrators, and practitioners. These different stakeholders tend not to share the same goals, and the result is an incoherent and disconnected education system that preserves through its division of labor a pronounced gap between the worlds of research and practice (Coburn & Stein, 2010; Penuel et al., 2011). Cobb and Smith (2008) emphasized that reformers must frame large-scale instructional improvement as a problem of school and district organizational learning that brings together researchers from different perspectives and extends what counts as worthwhile research.

THE ROLE OF RESEARCH FUNDING AGENCIES

Historically, research funding agencies may have been partially responsible for the problem of scale because they did not adequately account for the complexities of work in schools. Typically, funding agencies such as IES tested the effectiveness of interventions by using large-scale randomized controlled trials designed to evaluate the broad efficacy of intact, fully validated innovations. A principal research question in these studies, as noted in IES (2011) was whether an intervention worked “under different school and population conditions (e.g., urban vs. rural districts; with vs. without high proportions of English learners)” (p. 57). The intent was to determine whether the intervention was effective in spite of variations in context and population instead of looking for ways to account for these variations and build them into the research design. IES did not promote adapting or adjusting an intervention to fit new circumstances or studying the local context and the additional supports that might be necessary to promote implementation. However, the funding culture may be changing with a promising addition to the latest call for research proposals in which IES (2012) acknowledges the need to account for variations in context and population.

Another feature of IES-style scale-up evaluations has been that they require researchers to implement them under so-called typical conditions. For example, IES (2011) included the following:

Scale-up evaluations require that the intervention be implemented under conditions of routine practice. That is, the intervention should be implemented in the school or other authentic education setting as it would be if the school or entity had purchased and implemented the intervention on its own without any involvement in a research study... For Scale-up evaluations, the primary question of interest is, “Does the intervention produce a net positive increase in student learning and achievement relative to the comparison group under typical conditions?” (p. 56; italics in original)

Yet typical conditions have repeatedly been inadequate for supporting the sustained implementation of new interventions (e.g., Hitchcock, Dimino, Kurki, Wilkins, & Gersten, 2001). This way of thinking about scaling up appears to presume that interventions are a program or curriculum in a box with a unified set of procedures for implementers to follow. The assumption seems to be that teachers need only to adhere to a developer's directions with fidelity to reap the benefits of the program. However, as Cobb and Smith (2008) emphasized, it is not that simple; instructional practices are “complex, demanding, uncertain, and not reducible to predictable routines” (p. 5). These perspectives are at least partially why scale-up efforts have not fared better.

We propose an alternative way of thinking about scale-up research that includes ways to help districts build the capacity needed to support implementation of a new practice as an integral aspect of the research design. This focus on implementation draws from expertise in the
learning sciences. In their article on the implications of scale-up research for promoting the self-determination of individuals with developmental disabilities, Bacon and colleagues (2011) described implementation science as “the missing link that connects research outcomes to the delivery of effective practices” (p. 46). They noted that implementation science addresses critical issues that affect scale-up efforts, such as adoption decisions, capacity building, training, technical assistance, consumer participation and satisfaction, and long-term impact. Similarly, Odom (2009) portrayed implementation as the “link between evidence-based practices and positive outcomes” (p. 53).

Thus, instead of focusing only on whether an intervention is effective under typical practice conditions, we suggest that researchers ask questions such as the following:

- Under what conditions and with whom does the EBP work?
- What is necessary to support teachers’ implementation of the EBP?
- What is necessary to enhance the capacity of districts to support teachers in implementing an EBP under different ecological and population conditions?
- What is necessary to support broad, deep, sustained implementation of the EBP?

In the scale-up model that we recommend, researchers and practitioners work together to identify and examine the support structures that are most helpful in bringing about sustained use of an intervention (Cobb & Smith, 2008; Penuel et al., 2011). Researchers work side by side with teachers and district leaders to design, implement, and study the effects of these supports on classroom processes and student learning. In other words, instead of focusing exclusively on whether a particular innovation is effective, investigators strive to determine how to make the program work within a particular context (Bryk, 2009). Such research will likely require mixed methods that complement quantitative data with rich qualitative data that capture important contextual variables that influence successful scaling up (see Klingner & Boardman, 2011).

**Examples of Scale-Up Research in Special Education**

Special education researchers have used a variety of methodological approaches to focus on various aspects of scaling up. These scale-up efforts have been conducted in multiple schools in one district (e.g., using a variety of inclusive strategy approaches), in multiple states and districts (e.g., using Classwide Peer Tutoring [CWPT] and Peer-Assisted Learning Strategies [PALS]), and nationwide (e.g., using Schoolwide Positive Behavior Supports [SWPBS]). These studies shed light on some of the unique issues and challenges that efforts to scale up EBPs present.

Perhaps the most widely scaled-up practice has been SWPBS (Horner et al., 2009; McIntosh, Filter, Bennett, Ryan, & Sugai, 2010; Sugai, Horner, & McIntosh, 2008). As of 2008, nearly 8,000 schools in states across the country were at various stages of adopting and implementing SWPBS (Spaulding, Horner, May, & Vincent, 2008). Many of these schools have sustained implementation beyond 10 years (McIntosh, Horner, & Sugai, 2008); yet, in other sites, use has waned. Thus, SWPBS researchers are currently shifting their focus from establishing the efficacy of SWPBS on a large scale to documenting the processes that support systems change and identifying factors that promote scale-up and sustainability of EBPs (McIntosh et al., 2010). Such factors appear to include the following:

- Maximizing the contextual fit between the EBP and school needs.
- Promoting the EBP as a priority among implementers and stakeholders.
- Promoting effectiveness by ensuring fidelity of implementation.
- Increasing efficiency by integrating the EBP into daily school operations.
- Using data for continuous decision making to improve the fit, effectiveness, and efficiency of the EBP.

In another ongoing scale-up effort that spans 2 decades, Greenwood and colleagues have been exploring the use of technology as a way to support scaling and sustaining CWPT in multiple
states and districts (e.g., Greenwood, Delquadri, & Bulgren, 1993). Recently, Abbott, Greenwood, Buzhardt, and Tapia (2006) investigated the effectiveness and feasibility of a learning management system, online communications, and interactive multimedia resources for supporting CWPT implementation at several remote sites. Technology improved access and use of CWPT; however, the researchers determined that local conditions that support implementation are critical to promoting student achievement gains. Key elements of local support include strong local school leadership and the ability of a building faculty to work with information technology. Inconsistent communication or lack of timely communication among administrators, technical support providers, and teachers contributed to incomplete or no implementation of CWPT in a school.

In a related study, Buzhardt, Greenwood, Abbott, and Tapia (2006) noted that scale-up research generally lacks sensitive measures of the extent to which a practice is implemented with fidelity. They therefore examined how to assess the large-scale implementation of a practice in a technically sound and feasible manner. The researchers used distance communication technologies and a learning management system to monitor the rate of CWPT implementation across nine schools in five states for 1 school year. The implementation rate—or the number of weeks in which a school successfully completed 12 specific tasks needed to reach full implementation—varied substantially across schools (30 to 50 weeks). Buzhardt et al. identified communication challenges, limited on-site technical support, late implementation start-up, limited administrative support, and an overburdened site coordinator as barriers to scaling up.

In another multistate, multidistrict project, Fuchs and colleagues (e.g., Fuchs et al., 2010) attempted to scale up PALS in three sites: Tennessee, Minnesota, and South Texas. Their research focused on the level of professional development (PD) and ongoing support needed to scale up PALS. They first assigned kindergarten teachers randomly to a business-as-usual control group or to one of three levels of PALS support: (a) 1-day workshop, (b) workshop plus boosters (including two to three after-school problem-solving sessions with researchers and other PALS teachers), and (c) workshop plus booster plus helper (a member of the research team provided weekly technical support). Boosters appeared to add value to student reading outcomes, whereas using helpers to boost implementation fidelity did not (Fuchs et al., 2010), raising questions about the importance of implementing PALS with fidelity as opposed to allowing teachers some flexibility. Further, whereas PALS students outperformed controls on some beginning reading measures, PALS effects were strongest in the original research site (Nashville, Tennessee) and negligible in the most distal site (South Texas). Even in Nashville, the effects were not as pronounced as those found in previous efficacy research, most likely because of increased focus on implementing research-based reading instruction in business-as-usual conditions (Lemons, Fuchs, & Fuchs, 2008).

In response to these findings, the researchers changed their design in the latter part of the study to provide teachers with the option of customizing PALS. Teachers of Grades 2 to 5 implemented PALS for 1 year and then chose to continue either as “top-down” teachers, conducting PALS by the book, or become “bottom-up” teachers, with flexibility to customize noncore components of PALS. Fuchs et al. (2010) found that students in bottom-up PALS classes outperformed their peers in top-down PALS and control classes on reading measures. The researchers emphasized the importance of striking a balance between fidelity to core components of an EBP and flexibility to make the EBP meet particular student and classroom needs.

A smaller scaling-up study by Klingner, Ahwee, Piloniet, and Menendez (2003) focused on how to support the adoption of four research-based practices in inclusive classrooms across six elementary schools. These research-based practices were partner reading (from a version of CWPT; Mathes, Fuchs, Fuchs, Henley, & Sanders, 1994), Collaborative Strategic Reading (Klingner, Vaughn, & Schumm, 1998), Making Words (Cunningham & Cunningham, 1992), and Phonological Awareness Instruction (Torgesen & Bryant, 1994). Teachers participated in a 2-week summer workshop and received follow-up coaching throughout the school year. Teachers’ implementation of the practices varied. Klingner et al. (2003) sought to understand the barriers
and facilitators experienced by teachers who were high implementers ($M = 74.7$ implementations), moderate implementers ($M = 24.7$), and low implementers ($M = 3.3$). The top facilitator for high implementers was strong administrative support (as determined by analyzing teachers' logs and their responses to interview questions). Other facilitators that the high implementers described included support from university personnel, students' enthusiasm for the practices, and improved student achievement. Klingner et al. concluded that scaling up is not simply a matter of doing more of the same, but on a larger scale. For large scale implementation to occur, clearly there must be “buy in” at multiple levels. Unless reading leaders, district- and school-level administrators as well as teachers take over ownership of the practices, it is unlikely they will take hold and spread. This calls for a qualitatively different kind of involvement from researchers, from the higher levels where policy is decided to the classrooms where policy is carried out. (p. 427)

SCALING-UP CHALLENGES

Researchers face multiple challenges when scaling up EBPs, including differences between university and school cultures, varying perspectives on research, and educational policy considerations. When efforts to scale up interventions are unsuccessful or short-lived, the result is a gap between research and practice (Cobb & Smith, 2008; Odom, 2009; Penuel et al., 2011).

 Researchers and practitioners operate in different worlds, with distinct cultures, expectations, and hierarchies. They tend to interact almost exclusively with those in their respective communities to identify problems and plan solutions (Greenwood & Abbott, 2001). Lack of trust within and across organizations can thwart scale-up efforts. When researchers and practitioners attempt to work together, misunderstandings can easily occur.

School personnel do not always understand or value the knowledge generated by research. Not only can people with different beliefs interpret the same evidence in contrasting ways (Coburn, 2001), but there also is a strong tendency for district administrators (and others) to discount evidence when it does not support their preexisting beliefs or actions (Birkeland, Murphy-Graham, & Weiss, 2005; Coburn & Talbert, 2006). Levin (2010) reported that when he and colleagues asked school leaders about a series of research-supported propositions, the school leaders ranked personal experience and colleagues as more powerful influences on their beliefs than either PD or research in every case. Levin concluded that it is important to understand how district leaders find and use research and to understand the role of organizational factors in promoting the use of research.

Researchers and practitioners operate in different worlds, with distinct cultures, expectations, and hierarchies.

Similarly, Coburn, Toure, and Yamashita (2009) investigated the role of evidence in instructional decision making at the central office level. They found that district office administrators tended to search for and pay greater attention to evidence that resembled what they already knew and expected to find and that they did not necessarily even notice information that challenged their beliefs. They concluded that decision making in complex organizations like school districts is primarily based on interpretation, argumentation, and persuasion and that organizational and political factors—including the organizational structure of the district office, resource constraints, and leadership turnover—influence decision making.

Researchers may lament that administrators and teachers do not put enough stock in research, preferring instead to use approaches that are familiar and relying on personal beliefs and colleagues to determine which instructional practices to use (Vaughn et al., 2000). Teachers, however, might think that researchers operate in an ivory tower and are too disconnected from the realities that teachers face in schools, because researchers tend to interact primarily with other researchers.
and communicate their work primarily in scholarly articles and technical reports that are difficult for nonacademics to decipher (Boardman, Argüelles, Vaughn, Hughes, & Klingner, 2005). Teachers may sense that their expertise is undervalued (Klingner, Cramer, & Harry, 2006; McGill-Franzen, 2005). To address these challenges, Greenwood and Abbott (2001) emphasized the importance of researchers and practitioners working closely together to develop collaborative relationships.

A related challenge is that some researchers, program developers, and policy makers overstate the effectiveness of a practice and overgeneralize its applicability (McGill-Franzen, 2005). For example, Klingner et al. (2006) examined factors that contributed to the lack of successful scaling up of Success for All (SFA) in a large urban district (Urdegar, 2000) and found that teachers faced numerous challenges when trying to implement SFA by the book in situations different from those that developers of SFA anticipated. Teachers who drew on knowledge of their students and adapted SFA to be more appropriate for their circumstances seemed to apply the model most effectively. Klingner et al. (2006) cautioned against overgeneralizing the results of efficacy trials and recommended paying closer attention to issues of ecological and population validity. Like others, they called for a more nuanced look at context, as well as teachers' instructional decision making, when conducting scale-up research.

SCALING-UP SUPPORTS

Many factors potentially facilitate scale-up efforts. So that new practices will be sustainable, researchers and district leaders must integrate new practices with other district initiatives; and teachers should be able to see how the new practices facilitate their work rather than make it more burdensome. The new practices should not be tasks that teachers must perform in addition to everything else. This notion of integration suggests a need for new infrastructures that can support researcher–practitioner collaboration and lead to greater coherence (Penuel et al., 2011). District and research personnel should have a shared vision of effective instruction that guides their work, as well as a common discourse (Cobb & Smith, 2008). Ideally, brokers are in place who can connect various groups and bridge differing agendas (Fixsen & Blase, 2009). To foster greater coherence and collaboration between researchers and practitioners, all partners should make an effort to learn from and about one another’s goals for participation (Cobb & Smith, 2008). This ongoing effort requires refining and redefining roles over time, even when key personnel change. Researchers should learn about the processes that district leaders use in making decisions and attempt to understand how research is perceived as part of that process.

This idea that local contextual factors matter is prevalent in fields outside education. McDonald, Keesler, Kauffman, and Schneider (2006) wrote, “This simple notion—that it may be necessary to tailor an idea, product, process, or solution that ‘works’ to achieve consistently reliable results—is the hallmark of approaches to scale-up in marketing, manufacturing, strategic management, and medicine” (p. 16). They called for a tailored model approach to scale-up. For example, rather than best practices, which imply that educators should be able to apply solutions from one context to another, some experts in public policy recommend “smart practices” (Bardach, 2000). The basis for smart practices is the assumption that although educators can learn much from large-scale efficacy trials, applying what they have learned must take into account a number of variables specific to the context in which lessons are applied (Jabara & Dwivedi, 2004). Similarly, in medicine, clinicians draw from information on efficacy, possible adverse reactions, and contraindications in specific circumstances when determining treatment plans for individual patients. McDonald et al. noted that skilled teachers, like their counterparts in other fields, do not expect cookie-cutter solutions to suffice to address the challenges that they face but want and need more information about the essential features of an intervention and what they can adapt or adjust for changes in their environment and target populations.
PROFESSIONAL DEVELOPMENT IN THE CONTEXT OF SCALING UP

To promote such smart practices, effective PD is an essential ingredient in successful scale-up efforts (Chard, 2004; Desimone, Porter, Garet, Yoon, & Birman, 2002; Fogleman et al., 2006; Gersten, Chard, & Baker, 2000; Odom, 2009). Effective PD increases teachers’ knowledge, skills, and attitudes related to new practices, which in turn should lead to changes in instruction, which in turn should lead to improved student learning (Desimone, 2009). Traditional approaches to PD, which have typically entailed training teachers to implement new practices through brief one-time workshops, are generally insufficient for effecting meaningful, long-term changes (Birman, Desimone, Porter, & Garet, 2000; Garet, Porter, Desimone, Birman, & Yoon, 2001; Klingner, 2004).

Rather, more comprehensive and supportive PD systems are likely needed to ensure that teachers understand and can implement core components of new practices with fidelity (e.g., Kretlow & Bartholomew, 2010), adapt the practices to fit their specific contexts, and sustain them over time in real-world conditions (Pianta, Mashburn, Downer, Hamre, & Justice, 2008).

Odom (2009) referred to such comprehensive and supportive systems as “enlightened professional development” because they “recognize the necessity of steps beyond the workshop for taking EBPs into every day practice” (p. 59). Enlightened PD includes a strong focus on the social and organizational context in which schools and districts adopt new practices rather than merely encouraging individual teachers to implement new practices independently within their classrooms (Hutchings, n.d.). Enlightened PD approaches (cited by Odom, 2009) include team-building models, which involve various collaborative processes to adopt new practices; coaching and consultation models, in which support and feedback are provided from outside experts; communities of practice, in which teachers discuss and reflect on issues of common interest; and technology use such as online instruction and web-based interactive systems.

Regardless of the structure of the specific PD approach, Desimone (2009) outlined key features for effective PD (see also Desimone et al., 2002; Gersten et al., 2000; Klingner, 2004). These features include (a) a focus on specific content, including the conceptual underpinnings of that content; (b) active learning opportunities, such as through modeling, coaching, or discussing implementation efforts and problems; (c) extended opportunities for teachers to learn over time; (d) coherent learning opportunities that clarify how new practices align with school, district, and state standards and curricula; and (e) collective participation and collaborative opportunities through various structures, such as teacher study groups or grade-level teams. Researchers (e.g., Gersten, Dimino, Jayanthi, Kim, & Santoro, 2010; Klingner, 2004) have also emphasized the importance of the availability of data indicating that the practice has a positive impact on student achievement, as well as contextual supports, such as materials needed to implement the practice and administrative support.

In the context of scaling up a practice systemwide, the most productive PD efforts require substantial time and resources, as well as strategic and systematic planning by school leaders (Desimone, 2009; Odom, 2009). Researchers (e.g., Chard, 2004; Fogleman et al., 2006; Greenwood & Abbott, 2001) recommend that to maximize the success of PD models, districts and researchers should forge strong long-term partnerships, because they “provide an excellent test bed for developing strategies and resources for sustaining and scaling reforms so that they are more likely to become part of the district’s persistent institutional fabric” (Fogleman et al., 2006, p. 192). Strong partnerships require a commitment to building relationships and trust (Klingner, 2004) and can be mutually beneficial. For example, practitioners can benefit from ongoing support provided by the developers of new practices; and at the same time, researchers can learn much about how practices work in the real world, which can lead to development of even stronger innovations with greater potential for successful scaling up.

Administrative leadership and teacher buy-in are also important considerations for scaling up and sustaining EBPs (Kearns et al., 2010; Klingner, 2004). Approaches to cultivating strong administrative support and teacher buy-in include involving school and district personnel in identifying the instructional practices to adopt; engag-
ing administrators in discussing how to provide ongoing support to teachers; limiting the number of innovations introduced at one time; communicating the importance of the practices to teachers; providing systematic, ongoing feedback; identifying effective teachers and using them as models or mentors for new adopters; and providing time for planning and problem solving as well as opportunities to adjust and fine-tune implementation of new practices (Klingner, 2004). Emphasizing fidelity to core components of the practice, while allowing teachers flexibility to adapt some elements to suit their specific instructional contexts, may also increase the likelihood that teachers sustain new practices beyond the first year of implementation (Kearns et al., 2010).

Despite the consensus that successful scaling up depends on strong, ongoing PD, the field needs further research to help educators understand how best to provide this PD. In particular, empirical evidence of the effects of PD on student learning outcomes is sorely needed (Chard, 2004; Desimone, 2009; Gersten et al., 2010). Further, whereas it may seem logical to provide ongoing PD and support to promote teachers' adoption of scientifically based practices, such approaches can be extremely time-intensive and resource-intensive (Vaughn & Coleman, 2004). Intensive PD approaches may benefit teachers; but in the context of scaling up, they make little sense if educators cannot widely implement them. Important questions remain regarding how ongoing PD can support teachers' widespread use of EBPs in ways that promote student achievement and are feasible given the limited time and resources in schools.

Another area that is critical to creating and implementing PD in a scalable, sustainable way is preparing effective PD leaders (Borko, Koellner, & Jacobs, 2011). Research must not only develop strong PD models but should also provide insight into what it takes to implement PD in effective ways. Whereas traditional PD models have relied on trainers with expertise in the content that they will impart to teachers, enlightened PD models (Odom, 2009) will likely require experts not only in content but also in ways to facilitate active learning, community building, and sustained use of adopted practices.

**District Leadership That Supports Scaling Up**

The capacity of schools to successfully implement and sustain reform programs rests, in part, with district-level facilitation (Sanders, 2009). The role of the district is an important one. Yet, researchers tend not to pay enough attention to district partners. As noted by Rorrer, Skrla, and Scheurich (2008), “Intermittent attention to the district as the unit of study has left a void in our understanding of the complexities associated with the ability of district-level leaders to contribute to successful, systemic educational reform” (p. 307). Rorrer et al. proposed that there are four interdependent district roles: (a) providing instructional leadership, (b) reorienting the organization, (c) establishing policy coherence, and (d) maintaining an equity focus. Effective district leaders not only distribute leadership, but they also support the development of leadership capabilities. In other words, they build system capacity for extending and sustaining new initiatives (Cobb & Smith, 2008; Elmore, 2004).

For example, using distributive leadership (sometimes called shared leadership or team leadership; Spillane, 2005) as a guiding framework, Park and Datnow (2009) examined leadership practices in four urban school systems that were implementing data-driven decision making. The authors found that formal and informal leaders at school and district levels coconstructed data-driven decision making. Leaders played a pivotal role by helping positively frame the purpose of using data as a way to improve instruction. Leaders distributed decision-making authority in ways that built on the expertise and strengths of different staff members. Teachers learned to share student outcome data without fear of repercussion when scores were low, because the focus was on determining which instructional strategies were most effective, not on whether teachers were “good” or “bad.”

Johnson and Chrispeels (2010) investigated how linkages between a central office and district schools fostered professional accountability and organizational learning. Although teachers initially resisted the reforms, central office leaders, principals, and school leadership teams recognized the important role that PD could play in
supporting the district’s efforts to bring about change. They systematically looked for linkages, or ways to connect, with school personnel; and they built on the linkages to improve teaching and learning. One critical linkage turned out to be communication practices. The authors concluded that relational and ideological linkages are essential for enhancing commitment and professional accountability, as well as for ensuring a coherent instructional focus and organizational learning.

**AN EXAMPLE: COLLABORATIVE STRATEGIC READING IN A LARGE URBAN DISTRICT**

Collaborative Strategic Reading (CSR) is an EBP. Over a 12-year period, quasi-experimental and experimental research studies have found that CSR enhances reading comprehension in fourth-grade and middle school classrooms, with effect sizes ranging from .21 to .51 on the Gates MacGinitie Reading Test (2000), a group-administered standardized distal measure (Klingner, Vaughn, Argüelles, Hughes, & Ahwee, 2004; Klingner et al., 1998; Vaughn et al., 2011). However, CSR did not fare as well in a scale-up study in fifth-grade classrooms by Hitchcock et al. (2011), who found no statistically significant differences between CSR and comparison students on the Group Administered Assessment and Diagnostic Evaluation (American Guidance Service, 2001; a standardized distal reading test). Hitchcock and colleagues followed IES’s stipulation that an EBP should be scaled up under typical conditions of routine practice. The amount and quality of the PD and the support that teachers received did not appear adequate. Teachers received 2 days of initial PD and four coaching sessions (two during the fall and two during the spring). Although the coaches were experienced teachers and the investigator had trained them, they were new to CSR and none had taught it. Also, coaching sessions were limited to observations with feedback provided to the teachers and did not include planning before observations, coteaching, or modeling. Further, there were no follow up PD sessions to reinforce understanding of the CSR strategies or how to teach them to students. The researchers expected that teachers would learn everything that they needed to know in the 2 days of PD before implementation. The 37 CSR teachers were also spread across 26 schools, thereby limiting the potential for collaboration among teachers. The researchers checked fidelity only once per teacher, and it was low. The authors observed that only 21.6% of CSR teachers used all five core teacher strategies, which the authors defined as full procedural fidelity, whereas 56.8% of teachers used three or fewer strategies. Although CSR implementation may have been routine under these circumstances, it is unlikely that the PD and other forms of support that teachers received were sufficient to change their practice. Many teachers need a long time and a lot of support to learn how to teach comprehension strategies (Pressley & El-Dinary, 1997) and to implement cooperative learning (Klingner et al., 2004). We suggest that IES, researchers, and school districts may need to change their notion of routine practice to successfully implement interventions at scale. Does scaling up need to mean minimal support for teachers?

The following illustration of scaling up in practice during the first year of a school district-university-community partnership was part of a large Investing in Education (i3) funded validation grant, CSR-Colorado. We outline the process from planning to implementation and highlight three critical areas of coordination that support scaling up: (a) collaborating with district leadership, (b) coordinating PD and support structures, and (c) developing a shared understanding of research.

**COLLABORATING WITH DISTRICT LEADERSHIP**

CSR-Colorado has been a joint effort from the onset. Researchers and school district leaders wrote the research proposal in response to a request for proposals for a U.S. Department of Education i3 grant, CSR-Colorado. We began by collaboratively developing a shared vision for the project that related to the district’s needs and to the desired outcomes for school improvement:

[School District] is collaborating with [University] to validate an innovative, research-based intervention, Collaborative Strategic
Reading (CSR), in content area classrooms in eight middle schools. [School district] and its partners will use CSR to catalyze a whole school strategy to increase middle school teacher, principal, and staff effectiveness and academic achievement for English language learners (ELLs), students with learning disabilities (LD), and struggling readers. (Denver Public Schools, 2010)

This “broadcasted” vision statement encompassed the following two goals: (a) improve reading achievement for ELLs, students with LD, and struggling readers; and (b) use a whole-school strategy to do so. A third agreed-upon goal was to establish systems that would support scale up and sustainability for this as well as subsequent initiatives. Essentially, we were trying to change routine practice. The school district communicated its district messaging at board meetings, to the State Department of Education, and in PD at all levels, stating that the CSR project was a district initiative and that the district backed and supported it. The researchers also conducted numerous information sessions with various groups: state leaders, district coaches, area superintendents, and curriculum and program coordinators. We sought to cultivate the partnership between researchers and district personnel by establishing a culture of mutual respect. Outlining what each partner brought to the project and who was responsible for carrying out each component was an important step in the planning process. Figure 1 presents the initial areas of expertise that formed the foundation of our collaboration.

An accomplishment in the first year was our joint effort to establish roles, to learn from one another, and to distribute the work in ways that supported project goals and felt fair. As Park and Datnow (2009) indicated, collaborators should distribute decision-making authority among themselves to maximize the use of individual expertise. The partner who held the expertise for a particular project component was in charge of leading the work in that area, at least at the beginning. For example, as the developers of the CSR intervention, we researchers had expertise pertaining to the intervention itself, as well as an understanding of lessons learned over several years that related to PD and teacher support. Our role then was not only to provide PD, materials for implementation, and guidance about best practice but also to grow this knowledge within the district. Likewise, school district personnel understood the evolving and sometimes revolving systems and structures of the school district. They were aware of current and upcoming initiatives for teacher accountability (e.g., an overhaul to the teacher accountability system); student assessments (e.g., phase-out of state assessment, changing academic growth measures, and schedules for administration); curriculum (e.g., new standards); and PD (e.g., district-supported methods and structures for delivering PD, supporting teachers, and establishing the new teacher leader [TL] positions).
The district thus positioned itself to support such aspects of the project as integrating CSR with other district initiatives, identifying and bringing in school administrators, and aligning CSR PD and teacher support efforts with existing district practices.

An initial challenge for the research team pertained to learning to function in the very different culture of the district. Ways of doing, ways of talking, and ways of thinking about research all seemed very different from what we were accustomed to experiencing. We learned to expect frequent cancellations of meetings, sometimes without notice, and to understand that key participants would not always be present, because they were called to attend other meetings at the last minute. Follow-through was sometimes very slow, and there appeared to be little communication across some departments. (When the district hired new personnel to work on the project late in the first year, many of these issues disappeared.) We might have been able to minimize these frustrations, though seemingly minor in retrospect, if we had taken more time initially to set norms and establish and practice our roles.

**COORDINATING PD AND SUPPORT STRUCTURES**

The school district was in constant motion, with large initiatives moving in and out and with many initiatives in planning or trial phases; ours was one among many. From the start, we recognized that the success of the project would lie in our ability to coordinate with the school district to align CSR with existing structures and programs. Cobb and McClain (2001) termed these components (assessment practices, curriculum, district initiatives, etc.) the "institutional setting" of teaching. Whenever possible, the recommendation is to examine systematically the institutional setting before implementation and to adjust the research plan and design according to the findings (Cobb, Zhao, & Dean, 2009). School district employees, school administrators, and teachers needed CSR to integrate well into what they were already doing. We saw our PD and support structures as a vehicle for coordinating our efforts and as a key to successful scaling up.

The researcher’s lens often focuses at the classroom level, supporting teachers to improve practice on a classroom-by-classroom basis. In scale-up research, the lens must refocus to include larger district structures. As a team, we shared the vision of moving beyond teacher-by-teacher change to create systems that district personnel could reproduce but that would still be flexible enough to handle the needs of individual teachers and the variation that we saw across classrooms and schools. On the basis of effective PD research described previously, in the first year of a school’s CSR implementation, we provided a PD and support effort that included 2 initial days of professional development for all participating teachers, four to six after-school site-based booster sessions, and two or three coaching sessions each month (e.g., planning meetings, modeling, coteaching, written feedback) that we individualized according to teachers’ needs and preferences.

We understood that this intensive model was important in the first year, while teachers and the school district were new to CSR but that it would not be sustainable in the long term. One way in which we worked to align teacher support to promote scaling up was through a TL initiative, a district structure that was also in its first year of implementation in the district. Our plan included shared support in schools during the first year (i.e., weekly classroom coaching visits by the university coaches, mentoring TLs, and district coaches) so that over time, the district would build capacity to support teachers entirely with TLs and other district structures. This scaffolded coaching model promoted scaling up by gradually releasing responsibility to the schools and the school district. We reasoned that when external grant funding was no longer available to support the project, scaling up and sustainability efforts would need to come from within the district. The more the CSR project aligns with and incorporates itself into existing district structures, the more likely it will be to continue.

**DEVELOPING A SHARED UNDERSTANDING OF EBP**

CSR is now a district-adopted initiative. As part of our collaboration with the school district, we are conducting a study of support structures and
implementation, as well as of student outcomes. We have collected a variety of data, including student outcome measures, CSR fidelity observations, implementation logs from teachers and TLs, field notes, and teacher surveys. We quickly perceived tension between maintaining research standards in data analysis and dissemination and pressure from the district to change teacher practice and demonstrate increased student achievement.

Dissemination of data required coordination. Because of their expertise in this area, university researchers and the external evaluator (an outside evaluation team hired to collect and analyze portions of the project data) conducted rigorous data collection and analysis that school district personnel did not always understand or value in the same way.

The more the CSR project aligns with and incorporates itself into existing district structures, the more likely it will be to continue.

For example, after numerous conversations in which district personnel asked for data showing that CSR "works" so that they could share it with others (such as donors and parents) and we responded that we had already provided that information, it became clear that they were not interested in statistically significant differences between CSR and comparison conditions. Rather, they wanted to know if students who used CSR gained significantly in reading on the state's high-stakes test. Or, as they put it, "Does CSR help students become better readers?"

Another example resulted from the belief of district leaders that CSR highly aligned with the district's teacher accountability system and that a teacher implementing CSR with high quality would also receive strong ratings on the accountability system for that observation. At one point during the study, school district personnel reported that CSR worked better in science than in social studies, based on teacher evaluation observations. The district erroneously drew this conclusion from district accountability system observations of teachers who had or had not participated in the CSR PD, without accounting for whether a teacher was teaching CSR at the time of the observation. Once alerted to this discrepancy, the district agreed to describe the observation data more accurately. In addition, researchers are now working more closely with district leaders to ensure consistent assessment of teacher quality across observation measures. As university researchers, we needed to work very closely with our school district partners to establish how to describe and analyze CSR data and, more importantly, to communicate before disseminating any study results. This data snafu was an important red flag that demonstrated how quickly school personnel could draw conclusions about the effectiveness of CSR without fully understanding the research process.

**CONCLUSION**

It is unlikely that special education researchers will be able to solve the scale-up conundrum alone. Thus, special education researchers can benefit from collaboration with colleagues in the learning sciences, policy, and educational leadership. Similarly, special education researchers must learn to collaborate with their school district partners. We cannot overemphasize the importance of building relationships and becoming border crossers comfortable in different worlds—in universities, in schools, and in district offices. Yet relationship building occurs in an ever-changing, fluid context. Special education researchers need to determine how to build capacity and an institutional memory for their work. They must account for changes in context when scaling up and make sure that PD activities provide the support that teachers need for success over the long term.

Like colleagues in the learning sciences, we recommend a more nuanced approach to scale-up research that includes a plan for supporting teachers' implementation of an EBP. Design-based implementation research principles (Penuel et al. 2011) intrigue us; these principles emphasize (a) a focus on persistent problems of practice from multiple stakeholders' perspectives; (b) a commitment to iterative, collaborative design;
(c) a concern with developing theory related to both classroom learning and implementation through systematic inquiry; and (d) a concern with developing capacity for sustaining change in systems.

We also subscribe to the theory of action put forth by Cobb and Smith (2008). Following is a variation of their design principles:

- Establish nested learning communities.
- Specify instructional programs districtwide.
- Focus at every level (within the district) on classroom instruction, including core principles of learning and teaching.
- Establish two-way accountability in coordination with school and district staff.
- Promote the view that everyone is a learner.
- Align new practices with district standards, curriculum, assessment, and professional development.
- Provide continuing PD, based in the schools and linked to the instructional program.
- Create sustainable teacher support structures.
- Make pervasive use of data in decision making at all levels.

Scaling up EBPs presents a host of challenges to researchers and practitioners that go far beyond simply demonstrating the effectiveness of practices on a large scale. Certainly, researchers must conduct rigorous evaluations of EBPs taken to scale, but must do so in ways that account for the complexities associated with the processes that schools and districts use in adopting and sustaining new practices. To be successful, such efforts will likely involve long-term, collaborative partnerships with schools. Such efforts are of critical importance if EBPs are to have a positive, long-term impact on the students for whom they were designed.

REFERENCES


Cobb, P., Zhao, Q., & Dean, C. (2009). Conducting design experiments to support teachers’ learning: A


ABOUT THE AUTHORS

JANETTE K. KLINKNER (Colorado CEC), Professor; and ALISON G. BOARDMAN (Colorado CEC), Assistant Research Professor, School of Education, University of Colorado Boulder. KRISTEN L. MCMASTER (Minnesota CEC), Associate Professor, Department of Educational Psychology, University of Minnesota, Minneapolis.

Address correspondence concerning this article to Janette Klingner, School of Education, University of Colorado Boulder, 249 UCB, Boulder, CO 80309-0249; (e-mail: janette.klingner@colorado.edu).

The Collaborative Strategic Reading Colorado research project described in this article was supported by U. S. Department of Education Investing in Innovation (i3) Grant U396B100143.

Manuscript submitted April 2012; accepted July 2012.