Article

Educational Policy
2015, Vol. 29(1) 257–277

DOI: 10.1177/0895904814560886
epx.sagepub.com

Implementing Educational Innovations at Scale: Transforming Researchers Into Continuous Improvement Scientists

Lora Cohen-Vogel[1], Ariel Tichnor-Wagner[1], Danielle Allen[1], Christopher Harrison[1], Kirsten Kainz[1], Allison Rose Socol[1], and Qi Wang[1]

Abstract

There is growing concern among researchers and governmental officials that knowing what works in education is important, but not enough for school improvement. Sound evidence alone is not sufficient for large-scale, sustainable change, both because practitioners may consider it irrelevant to their own problems of practice or run into challenges when they try to implement. Failed attempts at replicating positive outcomes in new (or simply expanded) settings underscore the need for a different relationship between research and practice, one that takes a systemic perspective on improvement and transforms the role for research. In this article, we describe the new science of improvement and where it sits in the evolution of research on education policy implementation. We discuss the roots of the approach as well as its key features. We explain how the work differs from that of traditional research and end with illustrations of this difference from our experiences with the National Center on Scaling Up Effective Schools.

[1]University of North Carolina at Chapel Hill, USA

Corresponding Author:
Lora Cohen-Vogel, School of Education, University of North Carolina at Chapel Hill, Peabody Hall, Campus Box 3500, Chapel Hill, NC 27599-3500, USA.
Email: Lora.cohen-vogel@unc.edu

Keywords
continuous improvement, improvement science, education policy implementation, scale

> [The science of improvement] emphasizes multiple, small rapid tests of change by varied individuals working under different conditions. When this activity is organized around causal thinking that links hypothesized solutions to rigorous problem analysis and common data, we accelerate learning for improvement at scale.
>
> —Anthony Bryk, Carnegie Foundation for the Advancement of Teaching

There is growing concern among researchers and governmental officials that knowing what works in education is important, but not enough for school improvement. Sound evidence alone is not sufficient for large-scale, sustainable change, both because practitioners may consider it irrelevant to their own problems of practice or run into challenges when they try to implement (e.g., Coburn & Turner, 2011; Fishman, Penuel, Allen, & Cheng, 2013). Failed attempts at replicating positive outcomes in new (or simply expanded) settings underscore the need for a different relationship between research and practice, one that takes a systemic perspective on improvement and transforms the role for research.

Understanding this, the Institute of Education Sciences (IES) has launched a set of new grant initiatives to support efforts to build research–practice partnerships focused on developing and refining methods for answering *what works, for whom, and under what conditions*. These initiatives are intended to encourage researchers to explore processes by which schools in different contexts can continue to improve over time. The Carnegie Foundation, arguably one of the most influential of American grant-making foundations, is also using its substantial influence and funds to reshape our thinking around implementation by actively propagating a science of improvement.

In this article, we describe the new science of improvement and where it sits in the evolution of research on education policy implementation. We discuss the roots of the approach as well as its key features. We explain how the work differs from that of traditional research and end with illustrations of this difference from our experiences with the National Center on Scaling Up Effective Schools (NCSU).

Situating Improvement Science in Research on Education Policy Implementation

Much has changed since 1991, when Allan Odden described findings from the first waves of research on implementation in education. Back then, scholarship suggested that local governments had neither the will nor the capacity to implement programs initiated by higher level governments (Murphy, 1971; Pressman & Wildavsky, 1973). In fact, studies at the time showed that local officials sometimes used money allocated for new educational programming for purposes other than those for which the programs were designed, leading to regulations emphasizing compliance and the correct use of funds (Barro, 1974; Peterson, Rabe, & Wong, 1991).

The second wave of research, focusing on implementation *after* the initial start-up years, however, challenged the notion that programming initiated by higher levels of government would never be implemented (Odden, 1991). Instead, longitudinal studies of state and federal categorical aid programs in education repeatedly found that time, coupled with "mutual adaptation," or opportunities for educators to tailor programs to meet their local needs and circumstances, led to support for new program initiatives, the local capacity to run them, and, ultimately, the provision of services to targeted student populations (Berman & McLaughlin, 1975; Birman, Orland, Jung, Anson, & Garcia, 1987; Jung & Kirst, 1986; Sarason, 1982). As it turned out, it *was* possible to implement state and federal education reform at the local level.

However, as Odden (1991) wrote almost 25 years ago, "claiming that programs get implemented is not the same as claiming they are effective" (p. 8). Indeed, in the early 1980s, studies began to emerge that suggested that programs, even when fully implemented, were not solving the problems for which they were created. Although students who received extra services were fairing somewhat better than similar students who did not, the effects were often small and eroded over time (e.g., Baker & DeKanter, 1983; Kennedy, Birman, & Demaline, 1986). Wave 3 of research on education policy implementation then, emanating from arguments that efforts to develop rules to influence compliance had neglected program quality and impact, focused not only on how to get programs implemented but also on *how to get them to "work"* (Elmore & McLaughlin, 1983; Odden, 1991).

Arguably the longest-lasting and best-funded, Wave 3 was distinct from its predecessors in terms of the features of the policies examined and the approaches used to study them. Echoing Odden, Honig (2006) argued that—beginning in the 1980s—the shifting features of education reform strategies away from categorical programs and toward system-wide changes in curriculum and instruction influenced not only how research was conducted but also what we learned from it. Reforms in this period were aimed at issues at the very center of the schooling enterprise—what Ogawa (2009) called its "core technology." These issues included who should teach, what should be taught, and in what manner (Cohen-Vogel, 2011; Cohen-Vogel & Rutledge, 2009; Furhman, Clune, & Elmore, 1988; Hauptli & Cohen-Vogel, 2013; Osborne-Lampkin & Cohen-Vogel, 2014).

Studies during this period revealed that implementing programs of these types was highly complex and whether they "worked" depended on the people and places involved (Honig, 2006). McLaughlin (1991), for example, found that program success relied heavily on (a) the expertise of educators in the specific practices they needed to apply and (b) opportunities for them to collaborate with other program implementers as they tried out the practices (see also Anderson et al., 1987). Others showed that state agency leaders, like "street level bureaucrats" themselves, were consequential to implementation (e.g., Furhman et al., 1988). Place, too, became central to understanding program success (Honig, 2006). During this period—later referred to as the effective schools movement (e.g., Purkey & Smith, 1983)—researchers flocked to high-performing schools, asking "What conditions explained their performance?"

The key lesson from this third wave of implementation research in education was that program effectiveness, like its implementability, is the product of interactions between policies, people, and places—in short, the local context in which the program is tried (Honig, 2006). In the words of Means and Penuel (2005) then, the question for educational researchers is not simply what is implementable and what works, but instead "what works where, when, and for whom."

Despite these understandings—understandings that date back to at least as early as 2005—the federal government continued its focus on "what works," a focus that began in earnest with the passage of the Educational Sciences Reform Act (ESRA) in 2002 (Cohen-Vogel & Hunt, 2007; Schoenfeld, 2006b; Slavin, 2004). Specifically, the Wave 3 What Works focus was advanced by federal investments in building the capacity of the field to conduct randomized control trials (RCT). Since 2002, hundreds of millions of dollars have been invested in pre- and post-doctoral training programs aimed at building Fellows' capacity for using experimental and quasi-experimental designs to evaluate education programs, practices, and policies and on the development of the What Works Clearinghouse (WWC). The Clearinghouse, an on-line resource, evaluates the quality of research evidence on different programs, products, practices, and policies in education. "By focusing on the results from high-quality research," according to the WWC website, "we try to answer the question 'What works in education?'" To meet WWC's quality benchmarks, studies must have used a limited set of research designs.[1] The WWC's highest evaluations are reserved for studies using experimental designs. Considered by some to be the "gold standard" of research designs, experiments rely on random assignment of study participants to treatment and control groups, a method that maximizes the ability to generalize about program effects by distributing varying conditions and characteristics between the groups. The focus, therefore, *is not on the conditions and contexts that enable program success but on program success despite them*.

In addition to an emphasis on experimental designs, ESRA was almost entirely focused on the *production of* high-quality research rather than on its *translation for practice* (Cohen-Vogel, 2014). There was one notable exception. Five years after ESRA was adopted, the WWC began releasing practice guides that presented concrete recommendations to educators based on research findings (Dynarski, 2008). Despite this, calls for relevance continued. The General Accountability Office (2013), for example, noted in its recent review of IES that more could be done to communicate and disseminate research findings; in particular, it recommended that IES work to establish processes to assess whether educators are aware of research findings related to their own practice.

Early in this decade, IES began to respond. The response took two primary forms. First, IES began to invest in research to understand the conditions under which programs are successful in achieving their desired ends. In particular, they initiated two new grant competitions emphasizing research–practice partnerships and continuous improvement research, respectively.[2] Second, IES awarded a new Center for Knowledge Utilization[3] aimed at collecting information about the conditions under which research is used in schools and the factors that promote its use in practice.

For its part, the House of Representatives voted on May 8, 2014, to reauthorize IES. The bill, the Strengthening Education Through Research Act (SETRA), requires IES to ensure practitioners have a stronger voice in the research process, shift spending away from building state data systems to using them to improve practice, and fund research that examines the implementation of a particular policy or strategy and not just its impacts (Klein, 2014). Moreover, in contrast to the 2002 law, which emphasized "scientifically-based research," the bill emphasizes the need for "scientifically-valid" and "relevant" research, opening the way for research methodologies that capture the context in which programs are implemented. As of the final submission of this article, the Senate had not yet voted on the measure.

Along with federal actors, the Carnegie Foundation for the Advancement of Teaching is challenging educational researchers to develop and refine methods for answering what works, for whom, and under what conditions. In so doing, it is supporting a particular approach it calls "improvement science" (Bryk, 2009; Park, Hironaka, Carver, & Nordstrum, 2013). Improvement science refers to an approach for improving quality and productivity in diverse settings. In the words of Carnegie's President Tony Bryk and colleagues (2010), it brings researchers and educators together to answer, "what is the problem we're trying to solve, what is the change we're putting in place, and how will we know if that change is an improvement?" It does so by structuring the work around cycles of improvement, in which partners develop, test, and refine interventions designed to solve specified problems. In the remainder of the article, we focus on improvement science, poised to constitute what could be characterized as the fourth wave of research on education policy implementation.

What Is Continuous Improvement and From Where Did It Come?

So, what is this new science of improvement and why is it expected to facilitate implementation? Improvement science, or *continuous improvement* as it is increasingly being called in education, is an approach that involves multiple tests of small changes that can cumulatively result in larger, system change (Morris & Hiebert, 2011). As an applied science, it emphasizes innovation prototyping, rapid-cycle testing, and spread to generate learning about what changes, in which contexts, produce improvements.

Improvement science has its roots in health care and industry (Bheuyan & Baghel, 2005; Deming, 1993; Shewhart, 1931). The individual most widely identified with it is W. Edwards Deming, a 20th-century statistician who "championed the belief that statistical theory shows how mathematics, judgment, and substantive knowledge work together to the best advantage" (Mann, 1993, p. 3). Deming completed his PhD in mathematics and physics from Yale and, after honing his skills in government agencies and university departments, at the invitation of the U.S. Occupation authorities, began a series of consultations in Japan following World War II. He is cited as the engine behind the post-war revitalization of the Japanese economy, and much of his success is attributed to well-articulated methodologies for improvement (J. C. Anderson, Rungtusanatham, & Schreoder, 1994).

One such methodology is the *Plan, Do, Study, Act* (or PDSA) framework that Deming, along with his friend, Walter Shewhart, a statistician at Bell Laboratories, are credited with developing. PDSA, as will be discussed in more detail in the next section, places short inquiry cycles and the analysis of data in the center of the improvement agenda. In the context of Japanese workplaces, Deming combined these methodologies for improvement with culturally appropriate philosophies, known in Japan as *kaizen*, that emphasize teamwork and morale building as essential to meaningful improvement.

Donald Berwick, a pediatrician in Massachusetts, was one of the early and most influential adopters of continuous improvement from a discipline outside business. Berwick (1989) contrasted continuous improvement methods with traditional approaches that used monitoring and evaluation to seek out "bad apples" in the workforce and target them for remediation or removal. In contrast to a *bad apple* approach that focused improvement efforts on a tail of the competence distribution in the workforce, Berwick argued that "the theory of continuous improvement works because of the immense, quantitative power derived from shifting the entire curve of production upward even slightly, as compared with a focus on trimming the tails" (p. 54). These keen statistical and human insights led Berwick to co-found the Institute for Healthcare Improvement (IHI) in 1989. Since its inception, IHI has driven health care improvement research; today, its website contains links to more than 1,000 publications related to improvement in health care delivery and outcomes (Shortell, Bennett, & Byck, 1998).

To help explain what might be characterized as a surge in the use of the improvement approach not only in public health but also among multinational corporations, Donald Berwick points to Associates in Process Improvement (API) and its book—Berwick calls it a "classic"—*The Improvement Guide: A Practical Approach to Enhancing Organizational Performance*. The bestselling guide helps readers increase the rate and effectiveness of their improvement efforts across diverse settings with specified tools and illustrations from the public and private sectors (see Langley, Nolan, Norman, & Provost, 2009).

Thanks in part to *The Improvement Guide*, the use of improvement science is growing among educators and educational researchers. We are personally familiar with faculty who are teaching the book in their courses at schools of education at Harvard, Vanderbilt, and the University of North Carolina at Chapel Hill. Education groups such as the Carnegie Foundation and the Strategic Education Research Partnership (SERP) have worked to bring the approach to problems of practice in colleges, schools, and classrooms. Moreover, leading researchers in the learning sciences have successfully used designed-based implementation research, a close cousin of improvement science, smoothing the way for its adoption in education generally (Collins, Joseph, & Bielaczyc, 2004; Schoenfeld, 2006a). The National Society for the Study of Education released its yearbook on the topic late last year (see Fishman et al., 2013).

Key Features of the Approach

The Model for Improvement has two fundamental parts: (a) Three questions to guide the work—What are we trying to accomplish? How will we know that a change is an improvement? What changes can we make that will result in improvement?—and (b) the PDSA cycle to test changes in real settings (Langley et al., 2009). Guiding tests of change, the PDSA cycle helps

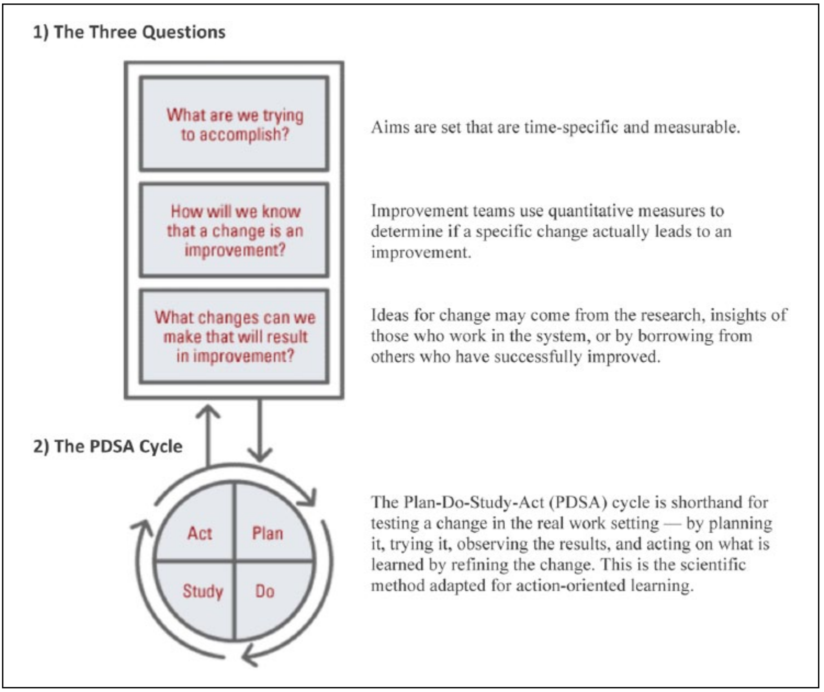

Figure 1. Model of improvement.
Source. Institute for Healthcare Improvement, 2014; Langley et al., 2009.
Note. PDSA = Plan-Do-Study-Act.

improvement teams determine whether a change is an improvement (Deming, 1993; Shewhart, 1939). The model of improvement is displayed graphically in Figure 1.

The PDSA cycle consists of four parts. First, the implementation team plans the test, asking what change ("prototype") will be tested, with whom/with what measures it will be tested, and what changes are expected as the result of trying out the prototype. Next, the team carries out the test, gathering information on what happened during the test and as a result of it. The team studies the information gathered during the test, comparing it with predictions made about the prototype's effects. Having studied the information, the team acts, making a decision about whether to abandon the prototype, revise it, or scale it up with a larger number of users.

After testing the change on a small scale—with a few teachers or classrooms—PDSA cycles repeat (see Figure 2). The improvement team learns from each test, refines the change, and then may implement the change on a

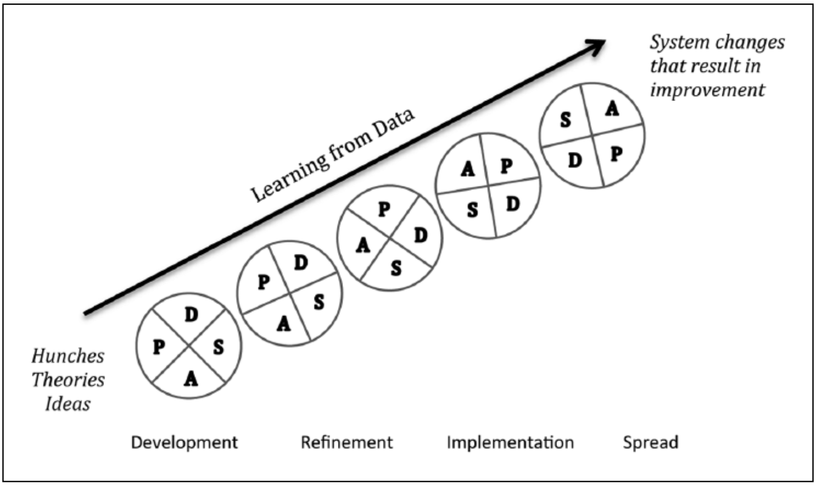

Figure 2. Repeated use of PDSA cycle for system change.
Source. Adapted from Deming, 2000; Langley et al., 2009.
Note. PDSA = Plan-Do-Study-Act.

broader scale—for example, with an entire grade level. After successful implementation within a unit, the team can continue to use PDSA to spread, or bring to scale, the change to other parts of the organization or other organizations entirely, effectively using it to adapt the change to new contexts, resulting in system change.

This kind of cycled inquiry can take various forms (e.g., RCTs, quasi-experimental designs) and affords adaptation to context or what some refer to as "local tests of change" (Park et al., 2013). Lengths of test cycles vary. Although short (90-day) cycles are common, improvement science allows for short- and long-cycle testing, depending on the research questions and nature of the change itself. By encouraging early and iterative testing of ideas in the specific environment of interest, the improvement model allows the innovation to be gradually modified to the uniqueness of the system in which it is being implemented (Langley et al., 2009).

In addition to the Model of Improvement, improvement science is concerned with developing capacity for sustaining change (improvements) in systems. In other words, an organization that conducts a one-time improvement project would not be said to be engaging in continuous improvement. Improvement science must be done with regularity and constancy. It must be fully integrated into the daily work of individuals within a system, be it a school, district, or state education agency (SEA; Park et al., 2013). Improvement science can facilitate system change through the development of organizational routines that help innovations travel through a system, habits of mind that conceive of teachers and other practitioners as co-creators in the design process, regular and ongoing assessments of organizational "readiness for change," improvement teams within school districts organized around persistent problems of practice, and participation in "networked improvement communities" or "hubs" with other school districts and researchers engaged in improvement initiatives (e.g., Bryk & Gomez, 2008; Cohen-Vogel, 2014; Fishman et al., 2013; Peurach & Glazer, 2012; Resnick & Spillane, 2006). In short, improvement science approaches should result in a self-sustaining learning process within an educational system.

Reaction Among the Educational Research Community

Reaction to the approach in education has been positive, with scholars and others speaking favorably about the direction IES is taking. Robert Granger, the immediate past president of the William T. Grant Foundation, for example, argues that IES's continuous improvement initiative could produce a new model for education research, where the gold standard "isn't to run a trial" but "seeing consistent results across a number of conditions" (as cited in Sparks, 2012). Stephen Raudenbush, chair of the University of Chicago's Committee on Education, agrees: "While we can learn an enormous amount from interventions, they won't [alone] produce the broad changes we need." For the Knowledge Alliance, Michele McLaughlin says improvement science is needed because "even the most cutting-edge practices, built on high-quality research and proven through rigorous testing, will have little measurable impact if not properly implemented" (as cited in Sparks, 2012).

To be clear, improvement science does not reject advances by IES, the Society for Research on Educational Effectiveness, and others to build capacity for research on causal effects of education practices, interventions, and policies (e.g., DeVinney, 2005). Indeed, W. Edwards Deming, generally considered the founder of improvement science, along with five of its key purveyors, the authors of *The Improvement Guide*, is a statistician and interested in the promise of planned experimentation (Langley et al., 2009). Instead, improvement science works to *combine* the power of investigations that have a strong basis for causal inference with subject area expertise, knowledge of design principles, and systems thinking to promote the translation, understanding, and use of evidence to improve decision making and outcomes (see Cobb & Smith, 2008; Penuel & Gallagher, 2009, for two examples).

An increasing number of researchers are using the approach. Apart from disseminating the approach among educational researchers, the Carnegie Foundation uses improvement science and is currently working with several school districts and community colleges across the country to develop, test, and refine innovations to solve significant problems of practice (see Carnegie's Statway, Quantway, Building a Teaching Effectiveness Network (BTEN) projects). The first author of this article is currently the co-principal investigator (PI) of the NCSU, a 5-year, $13.6 million initiative funded by IES that uses a continuous improvement model to bring to scale high school practices that have shown success in particular district contexts (see http://scalingupcenter.org). Another author is a leader with SERP. SERP uses design-based implementation research with public school systems in Baltimore, Boston, Oakland, and San Francisco, to iteratively develop and test innovations in varied settings to solve significant problems of practice.

There is evidence that the approach improves decision making and outcomes. In health care, there is a large literature establishing the efficacy of improvement models for depression care among veterans, diabetes care, and stroke care, for example (e.g., Benedetti, Flock, Pedersen, & Ahern, 2004; Power et al., 2014). A later adopter of the approach, the education field has a smaller but growing evidence base. Having begun in 2007, the Middle-School Mathematics and the Institutional Setting of Teaching (MIST) project provides what is perhaps some of the best evidence to date. Recognizing that the recent research in mathematics education—despite being rigorous—has had little influence on classroom instruction, MIST researchers worked with practitioners in four districts to establish an empirically grounded theory of action for improving the quality of mathematics instruction at scale (Cobb, Jackson, Smith, Sorum, & Henrick, 2013). The improvement approach they took led not only to new decision-making routines in the districts, wherein improvement strategies (e.g., providing teachers with access to a coach with instructional expertise in mathematics) were co-designed and refined in light of timely data about how they were playing out in schools, but also to robust instructional improvements in mathematics among teachers (Gibbons, Garrison, & Cobb, 2011; Smith et al., 2012). (See also the Center for Learning Technologies in Urban Schools [LeTUS]; the Next Generation Preschool Math project; and Carnegie's Teachings Pathways program.)

Improvement Science Versus Traditional Research

The improvement approach can be differentiated from traditional forms of research—and those from a quantitative paradigm, in particular—in several important ways.[4] First, whereas traditional research often aims to hold variables constant and uses a set of fixed procedures to carry out the work, improvement science "focuses on characterizing the situation in all its complexity" and uses an iterative, flexible process wherein design and research plans are revised as the work progresses (Barab & Squire, 2004, p. 4).

Second, whereas traditional research primarily focuses on outcomes of interventions, improvement science also *involves study of the design process*; there is an interest, that is to say, in understanding the workings of the approach itself in an effort to improve it the next time. Organizations, according to users, need "to improve their ability to improve" (Englebart, 1992, 2003).

Third, in improvement science, there is also a reconceptualization of the role of the researcher. Although pains are often taken in traditional quantitative research to keep researchers "outside" the intervention being tested, the improvement approach purposely *involves* the researchers in innovation design and revision. Moreover, researchers "are expected to become smarter about how to target issues that matter" to educators and about "how to conduct solid research within the constraints of practicing education systems" (Means & Harris, 2013, p. 360). The role of participants in the research is different too. Whereas traditional research often treats participants (e.g., teachers) as subjects, improvement science includes participants in the design process, involving them as equals in the work.

Illustrations of Researchers at Work in the Enterprise of Improvement Science

In this final section of the article, we describe each of these differences with illustrations from our work with the NCSU. Funded by the IES, the NCSU focuses on bringing to scale practices that make some high schools in large urban districts particularly effective with low-income students, minority students, and English language learners. In its first year, the NCSU conducted in-depth case studies in its two partner districts to identify the features that distinguished effective high schools from less effective high schools within them. In the Center's second phase, a District Innovation Design Team (DIDT) in both districts designed, developed, and tested innovations based on the Year 1 findings. Now, the DIDT, with support from the NCSU, is poised to implement the innovation in three innovation schools. To do so, it will continue its use of PDSA cycles to assess its implementation and proximate outcomes. As it does, it will follow a process for sharing successes and challenges with other implementing schools within the DIDT framework. By the conclusion of the Center's work, our aim is to have developed, implemented, and tested new processes that other districts will be able to use to scale up effective practices within the context of their own goals and unique circumstances.

Above, we argued that, unlike traditional quantitative research, improvement science focuses on characterizing the setting in all its complexity and uses an iterative, flexible process wherein design and research plans are revised as the work progresses. The NCSU approach highlights both of these differences. For example, in the first phase of research, the NCSU research team conducted comparative case studies of two higher and two lower value-added high schools[5] in each partner district. The purpose was to ascertain the components of effective high schools that distinguished them from less effective high schools with similar student demographics; how effective schools developed, implemented, integrated, and sustained those components; and the practices and policies that the effective schools used to leverage those components for successful student outcomes (Cohen-Vogel & Harrison, 2013). The research team initially designed data collection and analysis in each district using an a priori framework of eight essential components of effective high schools, identified through a comprehensive review of the research literature (Goldring, Porter, Murphy, Elliott, & Cravens, 2009). Not only did these case studies document the nuances of how each component was manifest in each school site, but the findings that emerged differed from the original eight components and differed *between* the two district contexts. In one district, *personalization for academic and social learning* distinguished the higher value-added from the lower value-added schools, while the distinguishing feature in the second district was *student ownership and responsibility for academic success* (Rutledge, Cohen-Vogel, Roberts, & Osborne-Lampkin, 2013). The subsequent plans for innovation design and development, along with the focus for future fieldwork, therefore, was adapted differently in the districts based on what worked in each context.

The design and development phase of the NCSU's research further highlights the ways in which improvement research uses an iterative and flexible, rather than fixed, process for research. Researchers, in collaboration with members of the DIDT and its school-level counterpart—the School Innovation Design Teams (SIDT)—developed "just-in-time" measures for PDSA cycles that tested the respective innovations. School-level personnel at each school site administered and analyzed these measures during each PDSA cycle. The short questionnaires, typically about five questions in length, provided valuable information in real time about the perceived effectiveness and feasibility of the innovation, which fed back into the design process as each school worked to adapt the innovation in ways that would work for their specific school. This illustrates that the process of improvement must use data, particularly data on outcomes, as it brings programs and practices to scale. Quick and easy-to-administer measures provide important information on what is working and what is not in any particular context, allowing site personnel to efficiently alter the program or practice in ways that will make the program or practice more likely to spread to other classrooms and/or schools and achieve the desired ends.

We also argued in the section titled "Improvement Science Versus Traditional Research" that improvement science approaches emphasize not only understanding the outcomes of educational interventions but also the workings of the approach itself in an effort to improve it the next time. As in traditional research, the NCSU collected quantitative data (e.g., school-wide teacher and student surveys; student test scores) and qualitative data (e.g., interviews with students and school personnel; classroom observations) to discern the extent to which the innovations in each district led to the desired outcomes. Yet, the NCSU's improvement research differed from traditional approaches in that multiple points of data were also collected on *the process* of developing, designing, and piloting the respective innovations. Members of the DIDT, including the research team, and SIDTs participated in cognitive interviews throughout the project to capture their understandings of the design and development processes and their suggestions for improvement. In addition, members of the research team completed reflection briefs after each bi-monthly DIDT meeting. During each of the DIDT and SIDT sessions, meeting participants completed feedback forms; data from each session was aggregated and used to plan subsequent sessions and track participants' understanding of and buy-in toward the improvement process and innovation over time. In addition, researchers observed and took field notes of DIDT and SIDT face-to-face sessions, summer institutes, and webinars. Data were synthesized through a process analysis, in which themes surrounding participants' attitudes, engagement, and understandings; the design challenge; and the design process, among others, were highlighted. Researchers were able to draw on the lessons learned from the data collection and analysis of the process itself to make recommendations for future improvement efforts (Harrison et al., 2014).

Finally, we argued above that, compared to traditional research approaches, improvement science changes the role of the researcher and study participants themselves. One key shift for researchers lies with a transition from the role of external investigator toward the role of participant in the work to design, implement, and improve educational innovations. For example, NCSU researchers, as members of the DIDTs, often found themselves acting as intermediaries between the world of educational research and their practitioner partners. In practice, this meant that researchers frequently attended design team meetings, helping to "translate" research—both our own and the relevant empirical research surrounding the ideas and designs articulated by the design team—for practitioners, and framing findings and key pieces of evidence in ways that enabled the team to engage in research-driven decision making. Furthermore, throughout the design process, the research team acted as full partners in the work—offering advice on key decisions, providing tools and resources on request, and taking an active role in facilitating the work of district partners as they planned for and executed the implementation of the innovation.

In addition, the NCSU research team worked to both complement and develop the capacity of our district partners as they utilized continuous improvement cycles to test, implement, and scale the designed innovation. Doing so successfully required the team to think carefully about balancing institutional norms of research and our partners' norms of practice. As we developed PDSA plans, for example, researchers and practitioners often engaged in the difficult process of creating instruments and data collection protocols that struck a balance between empirical rigor and the resource, time, and personnel constraints that practitioners faced in enacting them. The process also challenged the NCSU research team to alter its analytic processes to help district partners convert the data they collected into informative, accessible reports that were able to contribute to the iterative testing cycle in "just-in-time" fashion. Finally, both researchers and practitioners were often engaged as co-learners in the process, working together to adapt continuous improvement methods to the realities of the comprehensive high school environment.

Conclusion

As implied by these examples, improvement science approaches—approaches that seek to answer what programs work for whom and under what conditions—require a substantive shift in the role of researchers. It is a shift from the periphery of the process to its center. With that shift come conceptual and methodological challenges, as the roles between participants *in the process* and researchers *of the process* become blurred. Roles can become further blurred in the collaborations that form, as researchers work side by side with district officials, teachers, and a host of development specialists (e.g., curriculum developers; technology designers) to design, develop, and test educational innovations. As with other forms of participant research, researchers engaged in improvement science must carefully consider their own perspectives and positions within the improvement process at the same time they work to study it. Moreover, research teams engaged in improvement science must work to balance their own finite capacities, as they work to fulfill a twofold mission—to participate in and support the process, while simultaneously studying it.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305E100030. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Notes

1. See the study design standards at http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19
2. Researcher–Practitioner Partnerships in Education Research (Topic 1 under 84.305H); Continuous Improvement in Education Research (Topic 2 under 84.305H)
3. For more, see http://ies.ed.gov/funding/grantsearch/details.asp?ID=1466
4. Like Fishman, Penuel, Allen, and Cheng (2013), we argue that what is new about the approach is not in any one principle but in their integration.
5. See Sass (2012) for a description of the process used to identify and select higher and lower performing high schools for the study.

References

Anderson, B., Odden, A., Farrar, E., Fuhrman, S., Davis, A., Huddle, E., . . . Flakes-Mosqueda, P. (1987). State strategies to support local school improvement. *Science Communication*, *9*, 42-86.

Anderson, J. C., Rungtusanatham, M., & Schreoder, R. G. (1994). A theory of quality management underlying the Deming Management Method. *The Academy of Management*, *19*, 472-509.

Baker, K. A., & DeKanter, A. A. (1983). *Bilingual education*. Lexington, MA: Lexington Books.

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences*, *13*, 1-14.

Barro, S. M. (1974). *The impact of inter-governmental aid on public school spending* (Unpublished doctoral dissertation). Stanford University, CA.

Benedetti, R., Flock, B., Pedersen, S., & Ahern, M. (2004). Improved clinical outcomes for fee-for-service physician practices participating in a diabetes care collaborative. *Joint Commission Journal on Quality and Safety*, *30*(4), 187-194.

Berman, P., & McLaughlin, M. W. (1975). *Federal programs supporting educational change* (*Vols. 1-3*). Santa Monica, CA: RAND.
Berwick, D. M. (1989). Continuous improvement in an ideal in health care. *New England Journal of Medicine*, *320*, 53-56.
Bheuyan, N., & Baghel, A. (2005). An overview of continuous improvement: From the past to the present. *Management Decision*, *43*, 761-771.
Birman, B., Orland, M., Jung, R., Anson, R., & Garcia, G. (1987). *The current operations of Chapter 1 programs: Final report from the National Assessment of Chapter 1* (Report No. OR-87-504). Washington, DC: U.S. Department of Education. Retrieved from http://files.eric.ed.gov/fulltext/ED289935.pdf
Bryk, A. S. (2009). Support a science of performance improvement. *Phi Delta Kappan*, *90*, 597-600.
Bryk, A. S., Gomez, L. M., & Grunow, A. (2010). Getting Ideas Into Action: Building Networked Improvement Communities in Education, Carnegie Foundation for the Advancement of Teaching, Stanford, CA, retrieved from http://files.eric.ed.gov/fulltext/ED517575.pdf
Bryk, A. S., & Gomez, L. (2008). Ruminations on reinventing an R&D capacity. In F. Hess (Ed.), *The future of educational entrepreneurship: Possibilities of school reform* (pp. 181-206). Cambridge, MA: Harvard Education Press.
Cobb, P., Jackson, K., Smith, T., Sorum, M., & Henrick, E. (2013). Design research with educational systems: Investigating and supporting improvements in the quality of mathematics teaching and learning at scale. In B. J. Fishman, W. R. Penuel, A. R. Allen, & B. H. Cheng (Eds.), *Design-based implementation research: Theories, methods, and exemplars. National Society for the Study of Education Yearbook* (Vol. 112, pp. 320-349). New York, NY: Teachers College.
Cobb, P., & Smith, T. (2008). The challenge of scale: Designing schools and districts as learning organizations for instructional improvement in mathematics. In K. Krainer & T. Wood (Eds.), *International handbook of mathematics teacher education* (Vol. 3, pp. 231-254). Rotterdam, The Netherlands: Sense.
Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Perspectives*, *9*, 173-206.
Cohen-Vogel, L. (2011). Staffing to the test: Are today's school personnel practices evidence based? *Educational Evaluation and Policy Analysis*, *33*, 483-505.
Cohen-Vogel, L. (April, 2014). *Presidential invited session—Climbing out of the ivory tower: New forms of research–practice partnerships*. Presentation at the annual meeting of the American Educational Research Association, Philadelphia, PA.
Cohen-Vogel, L., & Harrison, C. (2013). Leading with data: Evidence from the National Center on Scaling Up Effective Schools. *Leadership and Policy in Schools*, *12*, 122-145.
Cohen-Vogel, L., & Hunt, H. (2007). Governing quality in teacher education: Deconstructing federal text and talk. *American Journal of Education*, *114*(1), 137-163/).
Cohen-Vogel, L., & Rutledge, S. (2009). The pushes and pulls of new localism: School-level instructional arrangements, instructional resources, and family–community

partnerships. In R. Crowson, & E. Goldring (Eds.), *The New Localism in American Education: The Yearbook of the National Society for the Study of Education*, Volume 108, (pp. 70-103). Chicago: University of Chicago Press.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences*, *13*, 15-42.

Deming, W. E. (1993). *The new economics*. Cambridge, MA: MIT Press.

DeVinney, B. (2005, September). *Expanding research and evaluation designs to improve the science base for health care and public health quality improvement*. Symposium at the Agency for Healthcare Research and Quality meeting, Washington, DC. Retrieved from http://www.hsrd.research.va.gov/quality2005/docs/summary.pdf

Dynarski, M. (2008). Bringing answers to educators: Guiding principles for research syntheses. *Educational Researcher*, *37*, 27-29.

Elmore, R. F., & McLaughlin, M. W. (1983). The federal role in education: Learning from experience. *Education and Urban Society*, *15*, 309-330.

Englebart, D. C. (1992). *Toward high-performance organizations: A strategic role for groupware*. San Jose, CA: Morgan Kaufmann.

Englebart, D. C. (2003, September). Improving our ability to improve: A call for investment in a New Future. In J. C. Spohrer, & D. McDavid (Chairs), *IMB Co-Evolution*. Symposium conducted at the meeting of the IMB Almaden Research Center. Retrieved from http://www.almaden.ibm.com/coevolution/pdf/engelbart_paper.pdf

Fishman, B., Penuel, W. R., Allen, A., & Cheng, B. H. (Eds.). (2013). Design-based implementation research: Theories, methods, and exemplars. *National Society for the Study of Education Yearbook*, Volume 112, Issue 2 (pp. 136 - 156). New York, NY: Teachers College Record.

Furhman, S., Clune, W., & Elmore, R. (1988). Research on education reform: Lessons on implementation of policy. *Teachers College Record*, *90*, 237-258.

General Accountability Office. (2013). *Preliminary observations on the Institute of Education Sciences' research and evaluation efforts* (GAO Publication No. GAO-13-852T). Retrieved from http://www.gao.gov/products/GAO-13-852T

Gibbons, L., Garrison, A., & Cobb, P. (2011, April). *Factors that influence coach centrality*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Goldring, E. B., Porter, A. C., Murphy, J., Elliott, S., & Cravens, X. (2009). Assessing learning-centered leadership: Connections to research, professional standards, and current practices. *Leadership and Policy in Schools*, *8*, 1-36.

Harrison, C., Redding, C., Brown, S., Wachen, J., Miller, J., & Socol, A. R. (2014). *Building capacity for improvement through collaboration: Findings from the National Center on Scaling Up Effective Schools*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Hauptli, M., & Cohen-Vogel, L. (2013). The federal role in education from Johnson through Obama: A policy regimes analysis. *American Journal of Education*, *119*(3), 373-404.

Honig, M. (2006). *New directions in education policy implementation: Confronting complexity*. Albany: State University of New York Press.

Institute for Healthcare Improvement. (2014). *Model for improvement*. Cambridge, MA. Retrieved from http://www.ihi.org/resources/Pages/HowtoImprove.

Jung, R., & Kirst, M. (1986). Beyond mutual adaptation, into the bully pulpit: Recent research on the federal role in education. *Educational Administration Quarterly*, *22*, 80-109.

Kennedy, M. M., Birman, B. F., & Demaline, R. E. (1986). *The effectiveness of Chapter 1 services: Second Interim report from the National Assessment of Chapter 1* (Report No. OR-87-501). Washington, DC: U.S. Department of Education. Retrieved from http://files.eric.ed.gov/fulltext/ED281919.pdf

Klein, A. (2014, May 8). Bill to make over federal education research quietly passes the House. *Education Week*. Retrieved from http://blogs.edweek.org/edweek/campaign-k-12/2014/05/bill_to_make_federal_education.html

Langley, G. L., Nolan, K. M., Norman, C. L., & Provost, L. P. (2009). *The improvement guide: A practical approach to enhancing organizational performance* (2nd ed.). San Francisco, CA: Jossey-Bass.

Mann, N. (1993). *Statisticians in history: W. Edwards Deming*. Retrieved from http://www.amstat.org/about/statisticiansinhistory/index.cfm?fuseaction=biosinfo&;BioID

McLaughlin, M. W. (1991). The Rand change agent study: Ten years later. In A. Odden (Ed.), *Education policy implementation* (pp. 143-155). Albany: State University of New York Press.

Means, B., & Harris, C. J. (2013). Towards an evidence framework for design-based implementation research. In B. J. Fishman, W. R. Penuel, A. R. Allen, & B. H. Cheng (Eds.), *Design based implementation research: Theories, methods, and exemplars. National Society for the Study of Education Yearbook* (Vol. 112, pp. 320-349). New York, NY: Teachers College.

Means, B., & Penuel, W. R. (2005). Research to support scaling up technology-based innovations. In C. Dede, J. Honan, & L. Peters (Eds.), *Scaling up success: Lessons from technology-based educational improvement* (pp. 176-197). New York, NY: Jossey-Bass.

Morris, A. K., & Hiebert, J. (2011). Creating shared instructional products: An alternative approach to improving teaching. *Educational Researcher*, *40*, 5-14.

Murphy, J. T. (1971). Title I of ESEA: The politics of implementing federal education reform. *Harvard Educational Review*, *41*(1), 35-63.

Odden, A. (Ed.). (1991). *Education policy implementation*. Albany: State University of New York Press.

Ogawa, R. T. (2009). Improvement or reinvention: Two policy approaches to school reform. In G. Sykes, B. Schneider, & D. N. Plank (Eds.), *Handbook of education policy research* (pp. 534-540). New York, NY: Routledge.

Osborne-Lampkin, L., & Cohen-Vogel, L. (2014). "Spreading the wealth": How principals use performance data to populate classrooms. *Leadership and Policy in Schools*, *13*, 188-208.

Park, S., Hironaka, S., Carver, P., & Nordstrum, L. (2013). *Continuous improvement in education*. Palo Alto, CA: Carnegie Foundation for the Advancement of Teaching.

Penuel, W. R., & Gallagher, L. P. (2009). Comparing three approaches to preparing teachers to teach for deep understanding in Earth science: Short-term impacts on teachers and teaching practice. *Journal of the Learning Sciences*, *18*, 461-508.

Peterson, P., Rabe, B., & Wong, K. (1991). The maturation of redistributive programs. In A. Odden (Ed.), *Education policy implementation* (pp. 65-80). Albany: State University of New York Press.

Peurach, D. J., & Glazer, J. L. (2012). Reconsidering replication: New perspectives on large scale school improvement. *Journal of Educational Change*, *13*, 155-190.

Power, M., Tyrrell, P. J., Rudd, A. G., Tully, M. P., Dalton, D., Marshall, M., . . .Parry, G. (2014). Did a quality improvement collaborative make stroke care better? A cluster randomized trial. *Implementation Science*, *9*(1), 40-49.

Pressman, J. L., & Wildavsky, A. (1973). *Implementation*. Berkeley: University of California Press.

Purkey, S., & Smith, M. (1983). The district policy implication of the effective schools literature. *The Elementary School Journal*, *85*, 353-388.

Resnick, L., & Spillane, J. (2006). From individual learning to organizational designs for learning. In L. Verschafeel, F. Dochy, M. Boekaerts, & S. Vosniadou (Eds.), *Instructional psychology: Past, present, and future trends* (pp. 259-279). Oxford, England: Pergamon.

Rutledge, S., Cohen-Vogel, L., Roberts, R., & Osborne-Lampkin, L. (April, 2013). *Understanding effective high schools: Findings and research around personalization for social and academic learning*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Sarason, S. B. (1982). *The culture of the school and the problem of change* (2nd ed.). Boston, MA: Allyn & Bacon.

Sass, R. (2012). *Selecting high- and low-performing high schools in Broward County, Florida for analysis and treatment* (Technical Report, National Center on Scaling Up Effective Schools). Available from http://scalingupcenter.org

Schoenfeld, A. H. (2006a). Design experiments. In P. B. Elmore, G. Camilli, & J. Green (Eds.), *Handbook of complementary methods in education research* (pp. 193-206). Washington, DC: American Educational Research Association.

Schoenfeld, A. H. (2006b). What doesn't work: The challenge and failure of the What Works Clearinghouse to conduct meaningful reviews of studies of Mathematics curricula. *Educational Researcher*, *35*(2), 13-21.

Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. Milwaukee, WI: American Society for Quality Control.

Shewhart, W. A. (1939). *Statistical method from the viewpoint of quality control*. New York, NY: Dover.

Shortell, S. M., Bennett, C. L., & Byck, G. R. (1998). Assessing the impact of continuous quality improvement on clinical practice: What it will take to accelerate progress. *Milbank Quarterly*, *76*, 593-624.

Slavin, R. (2004). Education research can and must address “what works” questions. *Educational Researcher*, *33*(1), 27-28.

Smith, T., Schmidt, R., Berebitsky, D., Garrison, A., Larbi-Cherif, A., & Cobb, P. (2012, April). *Relationships between school and district supports for adopting an inquiry-oriented curriculum and change in the quality of teaching*. Paper presented at the annual meeting of the American Educational Research Association, Vancouver, BC.

Sparks, S. D. (2012, October 17). IES to seed new methods for studying schools. *Education Week*, *32*(8), p. 1.

Author Biographies

Lora Cohen-Vogel is the Robena and Walter E. Hussman, Jr. Associate Distinguished Professor of Policy and Education Reform at the University of North Carolina at Chapel Hill (UNC). As the co-principal investigator of the National Center on Scaling Up Effective Schools (NCSU), she is interested in the political and organizational structures that facilitate or impede policy adoption, implementation, effectiveness and scale.

Ariel Tichnor-Wagner is a doctoral student in policy, leadership, and school improvement at UNC whose research areas include education policy implementation and teacher learning, with a specific focus on policies that improve outcomes for culturally and linguistically diverse students.

Danielle Allen is a doctoral student in policy, leadership, and school improvement at UNC whose research interests center on urban school reform efforts and education policy, with a critical eye toward school choice policies and barriers to access for students of color and students from low-income families.

Christopher Harrison is a doctoral student in policy, leadership, and school improvement at UNC whose work focuses on the implementation, scaling, and improvement of educational innovations and the politics of teacher evaluation and tenure reforms.

Kirsten Kainz, PhD, is the director of statistics at the Frank Porter Graham Child Development Institute and a research associate professor of education at the UNC. In addition, she serves as an education partnership consultant for the Strategic Education Research Partnership Institute in Washington, D.C.

Allison Rose Socol is a doctoral student in policy, leadership, and school improvement at UNC whose research interests include policies and practices in education that promote equity for underserved populations.

Qi Wang is a doctoral student in policy, leadership, and school improvement at UNC whose research interests include access to and equity of early education for children from low-income families, and using quantitative methods to study the impacts of educational policies.