2014 AERA Distinguished Lecture

Accelerating How We Learn to Improve

Anthony S. Bryk[1]

A chasm is growing between our rapidly rising aspirations for our educational systems and what schools can routinely accomplish. Education needs an improvement paradigm—one that recognizes the complexity of the work of education and the wide variability in outcomes that our systems currently produce. This article sketches out such a paradigm. It joins together the discipline of improvement science with the power of structured networked communities to accelerate learning to improve. These networked improvement communities (NICs) combine analytic thinking and systematic methods to develop and test changes that can achieve better outcomes more reliably. NICs are inclusive in drawing together the expertise of practitioners, researchers, designers, technologists, and many others. And they organize their activities in ways akin to a scientific community. They develop practice-based evidence as an essential complement to findings from other forms of educational research. The point is not just to know what can make things better or worse; it is to develop the know-how necessary to actually make things better.

Keywords: action research; collaboration; educational policy; educational reform; experimental design; improvement science; mixed-methods; organization theory/change; problem solving; research methodology; research utilization

This essay presents an argument about how educators might accelerate progress in their efforts to improve the educational opportunities afforded our nation's children.[1] I begin by setting a bit of context.

In late fall of 2007, I accepted the invitation from the Board of Trustees of the Carnegie Foundation for the Advancement of Teaching to become its ninth president. As part of the presidential search process, I shared with the board my views about the greatest challenges confronting the education field. For more than two decades, educators—practitioners, policymakers, researchers, and community leaders—have been actively pursuing reform. These efforts were manifest in a myriad of ways, including efforts to restructure the governance of public schooling, transform teacher education, develop comprehensive school reform models, and set more ambitious educational goals and accountability. Though these reforms were helping some schools improve, the problem was that our aspirations for schooling in America were increasing at a much faster rate. A chasm was growing between what we sought and what we could routinely achieve, and this chasm was greatest for our most disadvantaged students and communities. It seemed implausible to me that we could close this chasm absent a more vital R&D infrastructure focusing explicitly on improving teaching and learning in our nation's schools and strengthening the institutional arrangements in which educators and students carried out their work.

We want our schools to be more effective for many more students than ever before. Schools need to become more efficient in how they use their resources as constraints on public funding seem likely to persist. And far too many students continue to exit schooling absent adequate preparation for their lives ahead. Making meaningful headway on any one of these three aims would be a great accomplishment.[2] The call now was to advance simultaneously on all three. The search for a better way to address these extraordinary challenges has become the Carnegie Foundation's new mission.

[1]Carnegie Foundation for the Advancement of Teaching, Stanford, CA

This article was adapted/derived from the AERA Distinguished Lecture given on April 6, 2014, at the AERA Annual Meeting in Philadelphia, Pennsylvania. Portions of the article were subsequently elaborated in a book-length manuscript by Anthony S. Bryk et al. (2015). For more information, please visit http://hepg.org/hep-home/books/learning-to-improve. The present article was adapted with permission of the Harvard Education Press.

Educational Researcher, Vol. 44 No. 9, pp. 467–477
DOI: 10.3102/0013189X15621543

Education's Modus Operandi

Over the past several decades, many change efforts have moved rapidly across education but with little real knowledge as to how to effect the improvements envisioned by reform advocates (or even whether those improvements were possible). When reformers took aim, for example, at the high dropout rates and weak student engagement within high schools, a massive effort sprung forth to create new small high schools. Little guidance existed, however, as to exactly how to transform large, dysfunctional, comprehensive high schools into effective small schools. When reformers focused attention on weak in-service professional development, a whole new organizational role—the instructional coach—was introduced into schools (e.g., Elmore & Burney, 1997, 1998; Fink & Resnick 2001). What coaches actually needed to know and be able to do and the requisite organizational conditions necessary for them to carry out this work were left largely unspecified. When reformers recognized the importance of principal leadership, significant investments were directed at intensive principal development programs (e.g., Fink & Resnick, 2001). Principals were urged to become instructional leaders even though demands on their time were already excessive and few or no modifications were offered to relieve those demands. When policymakers were unsatisfied with the rate of school improvement, high-stakes accountability schemes were introduced. Unintended consequences abounded. The incidences of test score cheating accelerated, and select students were ignored as accountability schemes directed attention to some students but not others (e.g., Jacob & Levitt, 2003; Office of the Governor, 2011). The rapid introduction of value-added methods for assessing teachers began well before the statistical properties and limits of these methods were appropriately understood.[3] Not surprisingly, a host of problems has emerged, and political pushback has mounted.

In each instance, there was a real problem to solve, a nugget of a good idea, and in most cases, an academic research base that lent some credibility to the reform. Educators, however, typically did not know how to execute on these ideas. Districts and states lacked the individual expertise and organizational capacity to support these changes at scale. And many policymakers ignored arguably the most important factor for any of this to work: developing the will and agency among our nation's teachers and principals to engage productively with these reform efforts.

In general, the press to push good ideas into rapid large-scale use rarely delivers on the outcomes promised (for well-developed accounts on this point, see e.g., Fullan, 2001; Tyack & Cuban, 1995). At base is a common story of implementing fast and learning slow. As a field, we *undervalue the importance of systematic and organized methods of learning to improve*. When a pressing problem presents itself, we often jump to implement a policy or programmatic change before fully understanding the exact problem to be solved. We call this phenomena *solutionitis*. It is a form of groupthink in which a set of beliefs crystallizes based on an incomplete analysis of the problem to be addressed and without full consideration of potential problem-solving alternatives.[4] By seeing complex matters through a narrow lens, solutionitis often lures decision makers into unproductive strategies. With educational institutions under constant scrutiny, coupled with urgency for change, educators have become an easy target for such "solutions" to move quickly across our field.[5] We become disappointed when promised positive results do not readily emerge, and then we just move on to the next new idea. This should trouble all of us. If we continue to seek improvements in the ways we have always done, then we are likely to continue to get what we have always gotten.

Searching for a Better Way

A number of efforts seeking a more systematic approach to improving outcomes have emerged over the past decade in response to this concern. I review briefly three of these below.

Performance Management Thinking

Under this umbrella, specific measurable goals are articulated for each district, school, classroom, and different subpopulations of students. Data, typically in the form of standardized test scores, are used to chart progress toward these goals.[6] Educators are then held accountable for achieving these outcomes.

A positive aspect of this strategy is that more data are now available than ever before, and new tools have emerged that allow educators to examine these data in powerful ways. Educators are also now spending more time looking at such information, and greater transparency abounds about disparities in performance.[7]

Its weakness is that the strategy lacks an explicit theory about how educational practices might actually improve. Educators are told the targets they must achieve but then are left largely on their own to figure out how actually to hit these targets. If you thought that educators' lack of focus was the principle barrier to improvement, this approach might have real value. If, on the other hand, you view a lack of knowledge about how to achieve better student outcomes as the primary concern, this strategy has no obvious mechanism for generating such *know-how* and certainly not reliably at scale.

Evidence-Based Practice

Figure 1 illustrates a second and quite different approach to systematic improvement. This strategy draws impetus from clinical trials in medicine and seeks to transport such inquiries into education. Its strength is its emphasis on relevant theory and empirical rigor in drawing inferences about program effects. This is a much needed antidote in a domain where beliefs tend to run strong but supportive evidence is often weak or nonexistent.[8]

Primarily through the leadership of the Institute of Education Sciences, we have witnessed a significant expansion of such studies. Approximately 1,000 evidence-based entries now exist in the What Works Clearing House. To put this in perspective, some 27,000 clinical trials were reported on in medicine in just one year, 2010, and this inquiry base has been building for five plus decades. Yet even today, health care experts maintain that 80% to 90% of daily medical practice is not anchored in such evidence because the specific, detailed information practitioners need still does not exist (Institute of Medicine, 2012). While clinical trials are surely worth doing in education, they are a very slow and expensive process. From a purely practical point of view, such studies are not likely to be a primary resource for improving our schools anytime soon.

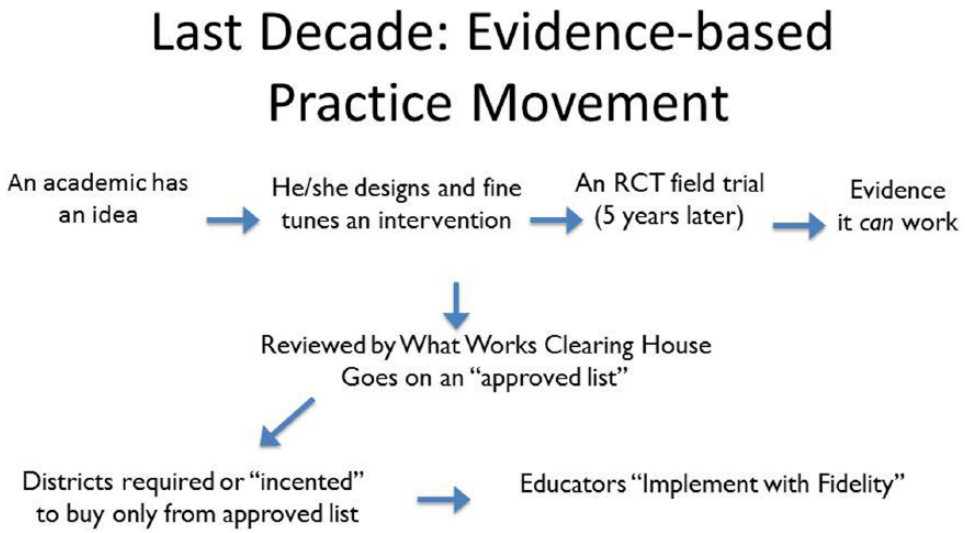

FIGURE 1. *The evidence-based practice strategy*

More fundamental, it is important to recognize that clinical trials target a narrow question. They tell us whether some intervention can work. That is, if a field trial produces a significant effect size, it means that the intervention had to work somewhere for somebody. Such studies, however, are not primarily designed to tell us what it will take to make the intervention work for different subgroups of students and teachers or across varied contexts. At base here is *the difference between knowledge that something can work and knowledge of how to actually make it work reliably over diverse contexts and populations*. Yet the latter is what practitioners typically want to know—what will it take to make it work for me, for my students, and in my circumstances? Unfortunately, policy actors who see evidence-based practice as today's answer typically miss this critical distinction.

Professional Learning Communities

The emergence of professional learning communities, also called communities of practice, represents a third alternative (for a seminal discussion of these ideas, see Lave & Wenger, 1991). These have become increasingly common in an effort to promote collaborative problem solving within schools and more broadly to connect educators who share common concerns. Unlike field trials that tend to focus on testing some new program or policy, these inquiries tend to zero in on very concrete day-to-day problems in the work of teaching and schooling. Educators are now cast as active agents of improvement rather than as passive receivers of knowledge developed by others. These communities seek to break down walls of solo practice and create safe spaces where faculty share and learn from one another.

Compared to clinical trials, the actual warrant for the exchanges that occur in these conversations remains less clear. Change ideas may not be well specified, with their success depending heavily on individual educators' tacit knowledge. Especially problematic for closing the quality chasm, there is no formal mechanism for accumulating, further detailing, and testing this individual clinical knowledge so that it might be transformed over time into a collectively held professional knowledge base (this theme is elaborated in Hiebert, Gallimore, & Stigler, 2002; Hiebert & Morris, 2009).

A Better Way: Improvement Science Within Networked Communities

Each of these briefly sketched strategies brings distinctive value, but each has its shortcomings as well. It is this observation that has impelled our work at the Carnegie Foundation for the Advancement of Teaching. Might there actually be a better way? A paradigm that keeps measureable improvements in valued student outcomes as its "north star." A paradigm that integrates extant research-based knowledge and draws on the analytic and empirical orientation of applied social science inquiry. A paradigm that also places primacy on addressing the specific questions that practicing educators confront and embraces their learning-by-doing orientation.[9] Equally important, a paradigm that sees educators as active inquirers who are now bound together by norms and structures akin to a scientific community. We have come to call this social organization a *networked improvement community* (NIC) and refer to its aims as building *practice-based evidence*.[10]

Every inquiry paradigm is anchored in some set of assumptions about the fundamental character of the phenomena we seek to learn about. Consequently, I detour briefly to consider two central features of schools as organizations: the complexity of the work we now ask educators to undertake and, in close tandem with this, the wide variability in performance that regularly results.[11] I argue that recognizing and wisely responding to these two issues—*systems complexity* and *variability in performance*—is key to any effort aimed at systematically improving the productivity of our educational institutions. The methods and organization of NICs take root here.

The Invisible Complexity in Our Educational Systems

For decades, medicine has developed extraordinary new technologies, life-saving drugs, and well-prepared clinicians. Atul

Gawande (2012b) tells us regarding his own specialization: "In surgery, you couldn't have people who are more specialized, and you couldn't have people who are better trained. Yet we still see unconscionable levels of death and disability that could be avoided." At base here is a modern-day conundrum for health care professionals.[12] Their field has more knowledge, more tools, and more resources than even before, but it remains seemingly unable to make all of this work reliably for advancing human betterment. Gawande emphasizes that "adding more parts" (meaning any combination of more people, new tools, and additional material resources), even great parts, does not assure a better result. Rather, he argues that we must attend to how all of this joins productively together for those charged with carrying out this work and for those that they seek to serve. In short, we must make service systems work better.

Reflecting on this quality chasm in medicine is instructive (Institute of Medicine, 2001). As noted previously, medicine benefits from great science, it invests heavily in developing its people, and it supports them with extraordinary technologies. Not surprisingly, educational reformers see all three of these factors as critical to improving the U.S. educational system. And indeed, we do need a stronger knowledge base, better professional education programs, and more effective use of technology to advance student learning. Yet, as in medicine, developments along these lines taken alone are not likely to redress the unsatisfactory student outcomes we now see.

The one-room schoolhouse was a relatively simple enterprise—a teacher instructing children of mixed ages, perhaps with chalkboards and some simple texts. Aspirations were modest: basic reading skills, some competency in arithmetic, and the ability to write one's name (for a more detailed historical description, see e.g., Graham, 2005). Now we expect students to master much more academic content. Literature texts previously taught in college courses, for example, have been pushed down into the high school grades. Students now study much more specialized topics in math and science courses than ever before. Moreover, we aspire that all students command not just basic skills but engage in deeper learning about the big ideas that organize our exploding knowledge universe (e.g., William and Flora Hewlett Foundation, 2010). These ever-expanding academic expectations greatly increase demands on teachers, both in their content knowledge and in their capabilities to teach.

Teachers' work also becomes more complex as classrooms grow more heterogeneous. We expect all teachers now, regardless of the subject taught, to be responsive to student cultures and languages different from their own. Along similar lines, reformers have advanced instructional initiatives, such as detracking, that make classrooms more heterogeneous in terms of students' backgrounds, abilities, and prior knowledge. This matters because the task of managing heterogeneous classrooms is a key teaching challenge (Cohen, 2011).

Likewise teachers, principals, and other local school leaders now interact with staff from a vast array of external service organizations. In seeking to close gaps in student achievement, many supplemental programs have sprung up, including individual tutoring, various forms of small group intervention, and after school and summer school programs. Instructional coaches specializing in areas such as mathematics, science, literacy, and student data analysis increasingly populate school buildings. Perhaps most demanding of all are the activities associated with the individual educational plans required by state and federal special education initiatives. The latter has brought an increasing array of specialists into schools—speech, hearing, learning disabilities, behavioral counselors, and so on. In principle, all of these efforts aim to support students' learning, but integrating them well poses a host of often unaddressed issues.

Especially significant, the density of this activity is greatest in our nations' most disadvantaged schools. These schools are doubly challenged in confronting the vast human needs of their student and family populations while at the same time having to manage a huge array of specialists, programs, and external connections targeting these students. The resultant increase in organizational complexity in these schools, however, tends to go unrecognized.[13]

Finally, educators confront a professional knowledge explosion. We are told today, for example, to attend to multiple forms of intelligence, differences in student learning styles, the salience of local communities, findings from brain research and cognitive science about how people learn, and from social-psychological on the importance of concepts such as resilience, grit, persistence, and student mindsets. Moreover, the list of new ideas continues to grow at a dizzying pace.

In sum, a number of forces have joined together to make the work of educators and the organization of schools more complex. As in health care, few professionals can effectively respond to all of this complexity in their daily work. So, it is not surprising that a chasm exists between what we seek to accomplish and what we actually achieve.

These observations suggest that we pay more attention to the tasks that educators take on and to the organizational environments that shape the way this work is carried out. Rather than believing that the route to improved outcomes is to continually add new programs (aka "more parts"), this perspective directs us to focus first on increasing our understanding of the work systems that are creating unsatisfactory results. For it is in this ability to *see the system* that meaningful progress often depends.

An example from the Baltimore Public Schools (BPS) illustrates how seeing this operational complexity can lead to a very different course of action. BPS was about to introduce a new process for evaluating teachers that included a protocol for principals to observe and rate the effectiveness of each teacher. In addition to informing personnel decisions, the district wanted to use the data to improve teaching. Typically, a central office team would have developed some new procedure for principals to provide teacher feedback based on these data. The district would then have conducted a training session for principals on the use of the protocol and directed them to implement it, several times a year, with each teacher.

Jarrod Bolte, formerly the director of teacher development for Baltimore Public Schools, had a different idea. Building on a basic improvement science principle—*be user-centered*—Bolte and his team asked a small group of new teachers to keep track of who visited with them to provide advice and feedback over a two-week period of time. As it turned out, some new teachers

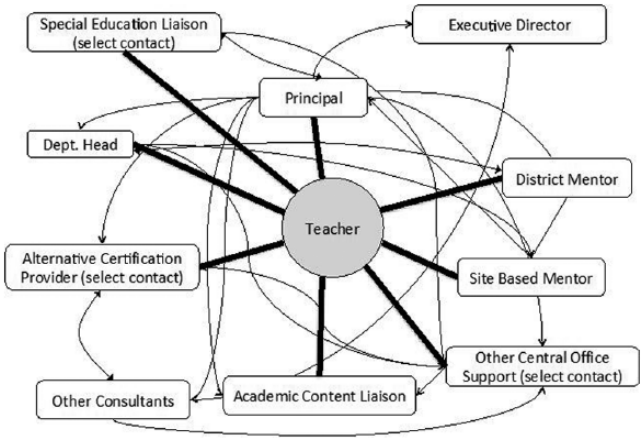

FIGURE 2. *Many new teachers experience a blizzard of guidance* Reprinted from *Learning to Improve: How America's Schools Can Get Better at Getting Better* (p. 29), by Bryk, Gomez, Grunow, and LeMahieu, 2015, Cambridge, MA: Harvard Education Press. Copyright 2015 by the President and Fellows of Harvard College. Reprinted with permission.

were already receiving lots of feedback—in some cases from 10 or more different sources (see Figure 2). At the same time, other new teachers received almost no feedback. So Baltimore had two quite different problems to solve. For some teachers, the issue was too much feedback from too many different people—feedback that was uncoordinated, often incoherent, and sometimes in outright conflict with one another. Elsewhere, the problem was not enough feedback. The challenge was how to assure that every new teacher received regular advice that might actually help him or her improve.

In short, the "aha moment" was recognizing that BPS had a complex and highly varied system of processes already at work that needed improvement. There were significant and unaddressed problems of coordination and communication as well as missed opportunities to enhance efficiency and effectiveness. Surely, the last thing one would want to do in this case is add still another disconnected process on top of all of this.

How does this happen? Clearly, no one consciously designed the system to function this way. Rather, it was the natural result of separate reform activities introduced over time by different institutional actors.[14] For example, state and district induction programs have mentors that regularly visit new teachers' classrooms. If the new teacher happened to be a Teach For America (TFA) corps member, he or she would also see a TFA coach every two weeks. (Other alternative teacher preparation programs and some colleges of education provide similar support.) This mentoring is layered on top of earlier reforms by the district's Office of Instruction, which introduced instructional coaches in literacy, math, and other subjects. The English Language Learner Office and the Special Education Department provided still more guidance. The Office of Instruction had also just introduced school-based learning cycles in which teachers were directed to use interim student test results to target immediate changes in instruction.[15] A new teacher may also have had an informal colleague or mentor. And all of this is in addition to the teachers' guides that accompanied the district's textbook series. (The latter can literally be a foot-high stack of materials.) Adding further to this mix is a vast array of professional support resources available through the district's online teacher portal and more generally across the Internet.

All of these supports were introduced with good intentions—to help teachers improve. No one, however, really attended to the overall work system that had been created here. More specifically, no one inquired about how all of this was actually being experienced by the teachers who it was supposed to help. In fairness, problems of this sort are not unique to education. For over 60 years, they have been grist for quality improvement efforts in other complex work settings. The achievements of improvement science in these other settings provide reasons to believe that these same principles and methods may help educators do the same.

Variation in Performance as the Problem to Solve

As American paleontologist Steven Jay Gould (1985) once wrote: "Variation itself is nature's only irreducible essence. Variation is the hard reality, not a set of imperfect measures for a central tendency." Gould's insight points toward a second driving concern guiding the efforts of networked improvement communities: Task and organizational complexity beget wide variability in performance. The resultant outcomes frequently resemble the familiar bell-shaped curve, illustrated in the left panel of Figure 3 (Gawande, 2012a). Improvement research seeks to reshape this distribution, moving overall performance to the right while dramatically reducing the frequency of negative outcomes. This is illustrated in the right panel of Figure 3.

Consequently, *understanding variation* in educational outcomes and responding effectively to it are the core goals of improvement research. Improvement science offers a very systematic but also highly practical set of principles and methods for advancing this learning. To highlight the need for such disciplined inquiries, we consider another educational example: a large-scale i3 field trial of the Reading Recovery program (May et al., 2013).

Reading Recovery targets first graders at risk for not learning to read. It is an intensive one-on-one tutorial that is anchored in over 30 years of research and practice. The program consists of about 50 sessions, each about 30 minutes in length, taught by a specially trained teacher. Figure 4 displays the results from Year 1 of the field trial (May et al., 2013). On average, the effect size is quite large, 0.70. But equally significant is the variability in these effects across schools. (Because the study was designed as a multisite trial, this variation is clearly visible. In cluster randomized trials, in contrast, it would be obscured.) Almost 20% of the schools had negative, null, or weak effects (less than 0.2). At the other tail of the distribution, another 20% of the schools had effect sizes of 1.2 or greater. Put somewhat differently, the variation in the effects across schools (i.e., the standard deviation among the school-level effects) is almost as large as the average effect.

Understanding the contours of this variation and the likely factors that contribute to it is key to achieving better outcomes more reliably at scale. The intervention clearly failed in a

FIGURE 3. *Variation in performance: what we typically get versus what we would like to have*
Reprinted from *Learning to Improve: How America's Schools Can Get Better at Getting Better (*p. 55), by Bryk Gomez, Grunow, and LeMahieu, 2015, Cambridge, MA: Harvard Education Press. Copyright 2015 by the President and Fellows of Harvard College. Reprinted with permission.

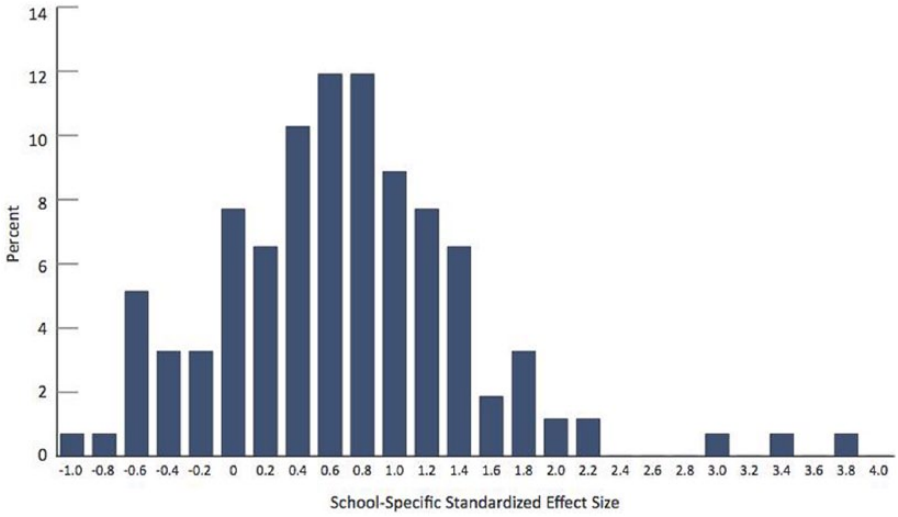

FIGURE 4. *Estimates of school-specific impacts of reading recovery*
Reprinted from *Evaluation of the i3 Scale-up of Reading Recovery: Year One Report, 2011–12* (p. 58), by May et al., 2013, Philadelphia, PA: Consortium for Policy Research in Education.

number of schools. We need to probe these problematic cases as to possible contributing factors. Might the program work better, for example, for some types of students but not others? Might variability in the preparation and/or skill of Reading Recovery teachers be a major concern? Did the program not integrate well with the core base of instruction in these schools? Or was there an operational breakdown in some schools so that the target students simply did not receive the normal course of the intervention? And if the latter, then why did this occur? Investigating these possible sources of variability—why null effects may be happening—can help direct where subsequent improvement efforts might focus. Such inquiries may also identify boundary conditions for the intervention—places (or subgroups of students) where it should not be used.

Equally interesting, very large effects accrued in other Reading Recovery sites. Might there be something we could learn from these positive deviants (Pascale, Sternin, & Sternin, 2010) that further strengthen the intervention for others in the future? In short, probing negative outcomes helps clarify problems still needing redress; examining positive deviants can help frame hypotheses about possible changes meriting future testing.[16]

It is important to recognize that each of the research methods we now commonly use originated to address a specific problem context. Herbert Weisberg (2014) in his book *Willful Ignorance* details the history of probability theory, including the emergence and use of the randomized trials in agriculture. The practical dilemma confronting agricultural experts was that many factors affected crop yield. It was difficult to tell whether, for example, the introduction of commercial fertilizer *added to* the overall (i.e., "on average") yield in the affected fields. The genius of randomization is that it allows researchers to discern the effect of adding some new factor, such as a fertilizer, willfully ignoring the influence of all the other factors that

contribute to the observed variability. However, when our concern shifts to how to achieve quality outcomes for every student and in every context, the details about "all the other factors" and how these factors may in turn *interact with* some new intervention really matter. Key here is the distinction between the words *interact* versus *add*. Interaction is a constitutive property of a system. Addition is not.

My point is straightforward: Randomized field trials are not principally designed to tell us how to make interventions work reliably for different subgroups of students and teachers working under varying contextual conditions. Improvement science, in contrast, places primacy on variability in outcomes as the central problem to address. It focuses our attention on how task and organization factors combine to create this variability.

In the most general terms, achieving quality outcomes reliably in complex systems, such as schools, requires its own distinctive methods of inquiry.[17] It favors the use of iterative testing to guide the development, revision, and continual fine-tuning of new tools, processes, work roles, and relationships. It also recognizes that we often do not know whether any change that we introduce will actually lead to improvement or if it might produce unintended negative consequences as well. Such ambiguity is intrinsic to complex adaptive systems.[18] So, this constitutive feature directs us toward a learning-by-doing orientation. It guides us toward starting small, learning quickly, and thereby minimizing the likelihood of harm in case we have it wrong. Change ideas are continuously refined based on evidence from what actually happens as the ideas are gradually taken up across increasingly diverse settings. Interestingly, we learn the most when a change fails to produce expected results or creates unintended consequences. This forces deeper thinking about the system we are trying to improve ("We must be missing something important here—what might it be?"). With subsequent cycles of redesign and testing, a better understanding evolves about the actual problem(s) to solve, and more productive changes are more likely to emerge. Consequently, it is sensible to think about these inquiries as cycles of *learning to improve*.

A Focus on Developing Practice-Based Evidence

As work becomes more complex, workers need to rely more heavily on good processes, well-designed materials, and explicit norms around communication (Hutchins, 1995). To the extent that at least select aspects of work can be formulated as routines, organizations create conditions that are more conducive to achieving quality outcomes reliably. Such routines free human attention to focus more sharply on details particular to the immediate situation.[19] Put differently: Absent good routines, highly variable consequences are the inevitable end result. This observation now grounds growing attention in teacher education research, for example, on detailing "high-leverage teaching practices" in specific content areas and grade levels.[20]

Likewise, as organizations become more specialized and complex, communication and coordination issues abound. The BPS efforts at improving teacher feedback, discussed earlier, illustrate this powerfully. When educational activities are poorly aligned, students may fall through the cracks as they move from one school to the next or from the general classroom to supplemental instruction. Communication breakdowns can also occur as students and teachers interact with counselors, social workers, family members, and others both inside and outside school. In confronting such organizational complexity, improvement research aims to synchronize the efforts of various individuals and organizations to reduce the chance of harm and improve the likelihood of better outcomes.

To be clear, responding effectively to issues of task and organizational complexity does not mean imposing some seemingly arbitrary standards from outside or delivered from above. Rather, it entails an education improvement community that actively creates them through its disciplined inquiries. Practitioners need to engage fully with researchers and others in developing, testing, and enhancing the clinical work of schooling.

Returning to Gawande's (2012b) observation about medical practice, securing better outcomes at scale means making "all of the parts" join more constructively together. How to make various work processes, tools, role relationships, and norms *interact more productively under a variety of conditions* is the improvement community's core learning goal. The resultant knowledge and associated empirical warrants are *practice-based evidence*. Improvement research, so focused, might help Reading Recovery, for example, to eliminate the occasional null and/or negative effects and concentrate its overall outcome distribution even further to the right.

Here is where the social organization of a networked improvement community comes to the fore. The many facets that can contribute to variability in educational outcomes (amid the task and organizational complexity sketched previously) transform improvement into a large-scale social learning problem. Moreover, many of the interventions that we seek to introduce into schools are themselves complex systems consisting of an interrelated set of principles, processes, tools, work roles, and norms. Consequently, advancing productive changes across diverse contexts entails coordinated, disciplined efforts among a large number of people and places. Recent developments in networked science illuminate how deliberately structured networks involving many different contributors can accelerate progress on complex problems (Nielsen, 2012). Participants in such networks join together around a shared working theory. They deploy common measures, inquiry methods, and communication mechanisms to anchor collective problem solving. They establish processes for how individuals carry out the work and establish evidentiary standards for warranting claims about effectiveness. Working together in these ways, they can learn faster how to improve.

Unfortunately, no professional infrastructure currently exists for educators to collaborate in the systematic development and testing of changes and to generate and synthesize practice-based evidence. But it could. Envision national networks of teachers and schools engaged with researchers and program developers around select high-leverage educational problems. These networks would aim to inform educators as to what is more likely to work where, for whom, and under what conditions. Moreover, as educators used this knowledge, the knowledge itself would evolve and be further refined through its applications.

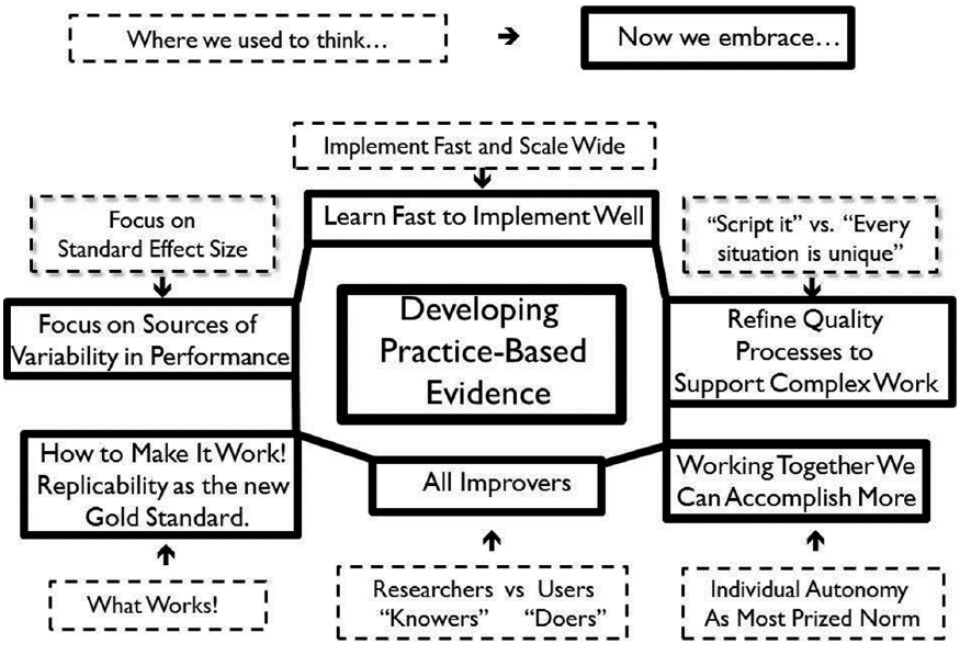

FIGURE 5. *The improvement paradigm*

Reprinted from *Learning to Improve: How America's Schools Can Get Better at Getting Better (*p. 186), by Bryk Gomez, Grunow, and LeMahieu, 2015, Cambridge, MA: Harvard Education Press. Copyright 2015 by the President and Fellows of Harvard College. Reprinted with permission.

An Improvement Paradigm

The notion of improvement research carried out through networked communities seeks to accelerate learning about the complex phenomena that generate unsatisfactory outcomes. This research activity forms around an integrated set of principles, methods, organizational norms, and structures. It constitutes a coherent set of ideas as to how practical inquiries should be thought about and carried out. Formally speaking, this is the definition of a paradigm. Once embraced, it calls us to think and act in very different ways. Figure 5 offers a heuristic guide for this improvement paradigm.

Where the urgency for change has in the past caused educational reformers to implement fast and scale widely—routinely underestimating the necessary know-how that needed to be developed and time required to succeed—an improvement paradigm directs us to learn fast in order to implement well under a variety of conditions and contexts. A continued failure to follow this dictum, based on experiences in numerous industries and sectors, will continue to relegate seemingly good ideas to the ever growing waste heap of promises unfulfilled (for an evocative example of this in the context of the military, see Sutton & Rao's [2014] discussion on "clusterfug," pp. 24–27).

Where researchers and analysts have traditionally focused on the precise estimation of an effect size—the average difference in outcomes between a group receiving an intervention and a control group that did not—improvement science directs us to scrutinize the myriad of factors that contribute to variation in outcomes and to use the insights gained to fuel continuous tests of change. The capacity to generate positive outcomes on a regular basis across varied conditions and contexts becomes the ultimate gold standard.

The improvement paradigm also represents a fresh way to resolve a long-standing dilemma for educational leadership, policy, and practice. The complexities embedded in teaching and learning have led many educators to believe that every situation is unique and that therefore, each educator, school, and district must invent practice on their own. While this view is grounded in an accurate appraisal of the intricacies of schooling, those who subscribe to this "solution" are essentially accepting that wide variability in outcomes at scale is inevitable. This fact has led others to advocate for scripting teaching so that every student might have the same opportunity to learn. But every student is not the same, nor is every context. The complexity is real, and it cannot be side-stepped by standardizing all activity in an effort to "teacher-proof" instructional environments.

In contrast to both of these views, an improvement paradigm recognizes task and organizational complexity as a central concern. It acknowledges the high demands placed on educators seeking to respond to each learner, the dynamics of each classroom, and distinctive features of each school community. An improvement paradigm also recognizes how important it is for educators to develop specific processes, tools, work roles, and relationships that scaffold quality work across these diverse settings. This approach respects the complexities involved in educating well while also affirming that important aspects of the work of schooling should follow standard protocols. Developing such practice-based evidence can help bridge the polarities described above.

Operationalizing improvement research also makes demands on what we measure and how we gather it. The field of education is now awash in data, mostly accountability data on students, teachers, schools, and districts. In contrast, improvement research calls for data not for purposes of ranking individuals or

organizations but for learning about how instructional practices and organizational processes actually work. We need evidence to help us discern whether any specific changes attempted are actually improvements. Introducing this kind of practical measurement poses a normative challenge as educators have traditionally seen data as intended for someone else—for a distant authority seeking to hold them accountable or a researcher studying them. Data have not been for educators to use to improve. Complicating matters further is the fact that time is the most restricted resource in education. A new challenge is how to harvest just enough information in the least obtrusive means possible to inform efforts to improve (for a further discussion of this idea, see Yeager, Bryk, Muhich, Hausman, & Morales, 2014).

Finally, an improvement paradigm calls us to reconsider our most prized educational norm: the autonomy of practice in classrooms and schools. In an earlier era, when we knew less, when teachers worked in simpler environments, and (perhaps most significantly) when our aspirations were more modest, it was reasonable to see the teacher as an individual craftsperson. Today, we have clearly reached the limits of what is achievable thinking of teaching in this way. Our problems are now too complex and the necessary knowledge and expertise too diverse. While some extraordinary individuals may well construct positive action on their own, left to this device, quality will always be in short supply. Fortunately, education engages hundreds of thousands of people doing very similar work every day. If educators joined together in structured improvement networks, our field would have extraordinary capacities to innovate, test, and rapidly spread effective practices (for a more detailed discussion, see Bryk, Gomez, & Grunow, 2010; Bryk, Gomez, Grunow, & LeMahieu, 2015). The improvement paradigm calls us to activate these strengths. It calls us to embrace a new narrative. Working together and learning together will make it possible to accomplish much more for many more than ever before.

Likewise, improvement science entails a different role for the education research community. The division between research and practice is legendary. We have tacitly accepted that there is a small class of "knowers" and a much larger class of "doers," who are expected to just use the knowledge generated by others. Improvement research, in contrast, calls on both to work together as "improvers." Respecting and valuing the varied expertise that is needed to solve educational problems, networked improvement communities embrace all involved as full members.

To sum up then: The improvement paradigm joins together the discipline of improvement science with the power of structured networked communities to accelerate learning to improve. It uses disciplined, analytic, and systematic methods to develop and test changes that achieve reliable improvements. It is inclusive in drawing together the expertise of practitioners, researchers, designers, technologists, and many others. And it is very deliberate in organizing its improvement activities in ways akin to a scientific community.

The paradigm sketched out here seeks to bring a valued but elusive goal—quality student outcomes reliably at scale—within our reach. It places emphasis on developing practice-based evidence as an essential complement to findings from other forms of educational research. The point is not just to know what can make things better or worse; it is to develop the know-how necessary to actually make things better.

Imagine a future in which this kind of systematic learning to improve is occurring every day in thousands of settings and engaging many thousands of educators, scholars, designers, and countless others. The education field could become an immense networked improvement community. We could greatly accelerate how we learn to improve. We could achieve valued outcomes that we now aspire to but realistically have no strategy to actually accomplish, whether it is all children reading by Grade 3, all children career and college ready by the end of high school, all children achieving a valued occupational certification, a two-year degree, or a four-year diploma—or all new teachers succeeding in educating their students well.

It is going to take a significant shift in our thinking and in the ways we work and learn together to get all of this to happen and to happen faster and better. I would welcome the opportunity to explore this challenge further with you and embark together as fellow travelers on the journey of learning to improve.

NOTES

[1]In attempting to reach a broad professional audience, I rely heavily on examples and the use of lay language. Although numerous citations and references are included, these too are offered primarily as examples of closely related work being advanced by many others. Given the breadth of issues introduced here and the text limits of an *Education Researcher* essay, no attempt has been made at a comprehensive literature review.

[2]We could also add to this list that professional morale is at an all-time low. See the Gallup (2014) *State of America's Schools* report. In a human and social resource intensive enterprise such as schooling, this too should be on our list of big problems to solve.

[3]See, for example, critical consensus reviews on value added methods at www.carnegieknowledgenetwork.org.

[4]This phenomenon is not peculiar to education but commonplace in other sectors as well. It is a natural tendency reflected in both individual and organizational actions. On this point, see the discussion in Pfeffer and Sutton (2000). See also Pentland (2014).

[5]This idea is supported in neo-institutional theory (see DiMaggio & Powell, 1983). In the absence of a broadly recognized professional status, institutions affirm their legitimacy in symbolic ways—by being on the cutting edge of reform. The new idea has currency precisely because it is new, regardless of what the problem actually is and the presence or absence of any evidence warranting the claims made.

[6]The annual yearly progress targets set out under No Child Left Behind offers a good known example of this.

[7]This may seem routine now, but it was not always the case. When we began the Consortium on Chicago School Research in 1990, no public data were regularly reported on disparities in achievement. Everyone knew these achievement gaps existed, but there was great ambivalence about explicitly reporting them.

[8]Examples abound here too. See, for example, the reinstitution of retention in grade initiatives even though multiple studies reported null or negative effects. On the latter, see the series of research reports by Melissa Roderick and colleagues at the Consortium on Chicago School Research, http://ccsr.uchicago.edu/publications/ending-social-promotion. Also see Roderick (1994).

[9]For a good introduction to this design-based orientation, see Dolle, Gomez, Russell, and Bryk (2014). For a more general discussion of design-based implementation research, see the edited volume by Fishman, Penuel, Allen, Cheng, and Sabelli (2013).

[10]The choice of words *practice-based evidence* is deliberate. We aim to signal a key difference in the relationship between inquiry and improvement as compared to that typically assumed in the more commonly used expression *evidence-based practice*. Implicit in the latter is that evidence of efficacy exists somewhere outside of local practice and practitioners should simply implement these evidence-based practices. Improvement research, in contrast, is an ongoing, local learning activity. We are indebted to Lawrence Green for introducing us to this concept of practice-based evidence and how it both relates to and contrasts with formulations about evidence-based practice. See Green (2008) and Barkham and Mellor-Clark (2003).

[11]This is recognized in the literature as a fundamental character of complex adaptive systems. See Axelrod and Cohen (2000) and Miller and Page (2007).

[12]Pondering this conundrum, Gawande (2012b) asks, "What if you built a car from the very best car parts? Well it would lead you to put in Porsche brakes, a Ferrari engine, a Volvo body, a BMW chassis. And you put it all together and what do you get? A very expensive pile of junk that does not go anywhere. And that is what medical practice can feel like today."

[13]See discussion about truly disadvantaged urban schools in Bryk, Sebring, Allensowrth, Luppescu, and Easton (2010, p. 209).

[14]This description is a generic account of a phenomenon that operates in large school districts. The specific details of this as experienced by an individual teacher will of course vary among both schools and districts.

[15]Many districts now deploy these interim tests, which are closely aligned to the end-of-year high-stakes accountability tests. While they are often described as formative assessments to improve instruction, some critics refer to these systems as "teaching to the test."

[16]Ironically, Reading Recovery is one of only a small number of educational interventions that had already been vetted and approved by the What Works Clearinghouse as having a strong evidence base. Consequently, the field probably did not need to spend $40 million (approximate cost of the field trial) to reaffirm this. Learning how to further improve Reading Recovery to achieve quality outcomes more reliably at scale might well have been worth a research investment. The latter would have cost less and probably learned more.

[17]For a full discussion of these principles and methods in the context of education, see Bryk et al. (2015); for a more general treatment, see also Langley et al. (2009).

[18]For further discussion of complex adaptive systems, see Axelrod and Cohen (2000) and Miller and Page (2007).

[19]Psychologists refer to this as automaticity. Such automaticity plays a key role in expertise development. It is a key feature distinguishing between novices and experts, and its development marks a key juncture in learning to move toward expertise. See Ericsson, Krampe, and Tesch-Romer (1993). See also Flyvbjerg (2001).

[20]This theme has been taken up recently in a variety of authors, including Ball and Forzani (2009); MacDonald, Kazemi, and Kavanaugh (2013); and Grossman, Hamerness, and McDonald (2009). The instructional practices of comprehensive literacy offer a well-detailed and concrete example. See, for example, Fountas and Pinnell (2001).

REFERENCES

Axelrod, R. M., & Cohen, M. D. (2000). *Harnessing complexity: Organizational implications of a scientific frontier*. New York, NY: Free Press.

Ball, D. L., & Forzani, F. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education*, *60*(5), 497–511.

Barkham, M., & Mellor-Clark, J. (2003). Bridging evidence-based practice and practice-based evidence: Developing a rigorous and relevant knowledge for the psychological therapies. *Clinical Psychology & Psychotherapy*, *10*(6), 319–327.

Bryk, A. S., Gomez, L. M., & Grunow, A. (2010). *Getting ideas into action: Building networked improvement communities in education*. Stanford, CA: Carnegie Foundation for the Advancement of Teaching.

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press.

Cohen, D. K. (2011). *Teaching and its predicaments*. Cambridge, MA: Harvard University Press.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*, 147–160.

Dolle, J. R., Gomez, L. M., Russell, J. L., & Bryk, A. S. (2014). More than a network: Building professional communities for educational improvement. In B. J. Fishman & W. R. Penuel (Eds.), *Design-based implementation research: Theories, methods, and exemplars* (pp. 443–463). New York, NY: Teachers College Record.

Elmore, R., & Burney, D. (1997). *Investing in teacher learning: Staff development and instructional improvement in community school district #2*. New York, NY: National Commission on Teaching & America's Future & the Consortium for Policy Research in Education.

Elmore, R., & Burney, D. (1998). *Continuous improvement in community school district #2*. New York, NY: University of Pittsburgh, Learning Research and Development Center, High Performance Learning Communities Project.

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 393–394.

Fink, E., & Resnick, L. B. (2001). Developing principals as instructional leaders. *Phi Delta Kappan*, *82*(8), 598–606.

Fishman, B. J., Penuel, W. R., Allen, A.-R., Cheng, B. H., & Sabelli, N. (2013). Design-based implementation research: An emerging model for transforming the relationship of research and practice. In B. J. Fishman & W. R. Penuel (Eds.), *National Society for the Study of Education* (pp. 136–156). New York, NY: Teachers College Record.

Flyvbjerg, B. (2001). *Making social science matter: Why social inquiry fails and how it can succeed again*. New York, NY: Cambridge University Press.

Fountas, I. C., & Pinell, G. S. (2001). *Guiding readers and writers (Grades 3-6): Teaching, comprehension, genre, and content literacy*. Portsmouth, NH: Heinemann.

Fullan, M. (2001). *Leading in a culture of change*. San Francisco, CA: Jossey-Bass.

Gallup. (2014). *State of America's schools: The path to winning again in education*. Washington, DC: Gallup.

Gawande, A. (2012a). The difference between coaching and teaching [Video file]. Retrieved from https://www.youtube.com/watch?v=VabtGPVVihA

Gawande, A. (2012b). How do we heal medicine [Video file]. Retrieved from http://www.ted.com/talks/atul_gawande_how_do_we_heal_medicine?language=en#t-3268

Gould, S. J. (1985). The median isn't the message. *Discover*, *6*, 40–42.

Graham, P. (2005). *Schooling America: How the public schools meet the nation's changing needs*. Oxford: Oxford University Press.

Green, L. W. (2008). Making research relevant: If it is an evidence-based practice, where's the practice-based evidence? *Family Practice, 25*(Supplement 1), I20–I24.

Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching: Theory and Practice*, *15*(2), 273–290.

Hiebert, J., Gallimore, R., & Stigler, J. W. (2002). A knowledge base for the teaching profession: What would it look like, and how can we get one? *Educational Researcher*, *31*, 3–15.

Hiebert, J., & Morris, A. K. (2009). Building a knowledge base for teacher education: An experience in K–8 mathematics teacher preparation. *Elementary School Journal*, *109*(5), 475–490.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, *19*(3), 265–405.

Institute of Medicine, Committee on Quality of Health Care in America. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: National Academies Press.

Institute of Medicine, Committee on Quality of Health Care in America. (2012). *Best care at lower costs: The path to continuously learning health care in America*. Washington, DC: National Academies Press.

Jacob, B. A., & Levitt, S. (2003). *Catching cheating teachers: The results of an unusual experiment in implementing theory* (Working Paper No. 9413). Retrieved from http://www.nber.org/papers/w9414

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

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York, NY: Cambridge University Press.

May, H., Gray, A., Gillespie, J., Sirinides, P., Sam, C., Goldsworthy, H., . . . Tognotta, N. (2013). *Evaluation of the i3 scale-up of Reading Recovery: Year one report, 2011–12* (CPRE Research Report No. RR-76). Philadelphia, PA: Consortium for Policy Research in Education.

McDonald, M., Kazemi, E., & Kavanagh, S. S. (2013). Core practices and pedagogies of teacher education: A call for a common language and collective activity. *Journal of Teacher Education*, *64*(4), 378–386.

Miller, J. H., & Page, S. E. (2007). *Complex adaptive systems: An introduction to computational models of social life*. Princeton, NJ: Princeton University Press.

Nielsen, M. A. (2012). *Reinventing discovery: The new era of networked science*. Princeton, NJ: Princeton University Press.

Office of the Governor. (2011, July 5). *Deal releases findings of Atlanta school probe* [Press release]. Retrieved from http://gov.georgia.gov/press-releases/2011-07-05/deal-releases-findings-atlanta-school-probe

Pascale, R., Sternin, J., & Sternin, M. (2010). *The power of positive deviance: How unlikely innovators solve the world's toughest problems*. Boston, MA: Harvard Business Review Press.

Pentland, A. (2014). *Social physics: How good ideas spread—the lessons learned from a new science*. New York: Penguin Press.

Pfeffer, J., & Sutton, R. I. (2000). *The knowing-doing gap: How smart companies turn knowledge into action*. Boston, MA: Harvard Business School Press.

Roderick, M. (1994). Grade retention and school dropout: Investigating the association. *American Educational Research Journal*, *31*, 729–759.

Sutton, R. I., & Rao, H. (2014). *Scaling up excellence: Getting to more without settling for less*. New York, NY: Crown Business.

Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.

Weisberg, H. I. (2014). *Willful ignorance: The mismeasure of uncertainty*. Hoboken, NJ: John Wiley & Sons.

The William and Flora Hewlett Foundation. (2010). *Education program strategic plan*. Retrieved from http://www.hewlett.org/uploads/documents/Education_Strategic_Plan_2010.pdf

Yeager, D. H., Bryk, A. S., Muhich, J., Hausman, H., & Morales, L. (2014). *Practical measurement* (Working paper). Retrieved from http://www.carnegiefoundation.org/wp-content/uploads/2014/09/Practical_Measurement_Yeager-Bryk1.pdf

AUTHOR

ANTHONY S. BRYK, EdD, is the ninth president of the Carnegie Foundation for the Advancement of Teaching, 51 Vista Lane, Stanford, CA 94305; *t.bryk@carnegiefoundation.org*. He is leading work on transforming educational research and development, more closely joining researchers and practitioners to improve teaching and learning.

Manuscript received November 5, 2015
Accepted November 17, 2015