What Is Improvement Science? Do We Need It in Education?

Catherine Lewis[1]

The theory and tools of "improvement science" have produced performance improvements in many organizational sectors. This essay describes improvement science and explores its potential and challenges within education. Potential contributions include attention to the knowledge-building and motivational systems within schools, strategies for learning from variations in practice, and focus on improvement (rather than on program adoption). Two examples of improvement science in education are examined: the Community College Pathways Networked Improvement Community and lesson study in Japan. To support improvement science use, we need to recognize the different affordances of experimental and improvement science, the varied types of knowledge that can be generalized, the value of practical measurement, and the feasibility of learning across boundaries.

Keywords: organization theory/change; professional development; teacher education/development; mathematics education

The failure of research-based knowledge to "scale up" broadly is a central challenge in education (Coburn & Stein, 2010). As one U.S. education school dean wrote, "why is so much that is known about how to help U.S. students reach high levels of achievement not applied in most school settings?" (Stipek, cited in Coburn & Stein, 2010, p. xi). New approaches to the connection between knowledge and improvement are emerging from improvement science (Langley et al., 2009), an applied science that has dramatically improved practice in industries ranging from automobile manufacturing (Rother, 2009) to health care (Gawande, 2007).[1] As projects rooted in improvement science begin to show success within education (Bryk, Gomez, Grunow, & LeMahieu, 2015), the moment is right to consider its underlying tenets.

What Is Improvement Science?

Imagine that a district wants to improve students' capacity to "make sense of problems and persevere in solving them" (Common Core State Standards Initiative, 2010, p. 6). Improvement science theorizes that two different types of knowledge are needed: basic knowledge from the discipline of education (for example, knowledge about effective mathematical tasks and instructional strategies) and "a system of profound knowledge" needed to *enact* basic disciplinary knowledge within organizations (Deming, cited in Langley et al., 2009, p. 75). The "system of profound knowledge" is drawn from sociology, psychology, and statistics and includes "knowledge of systems, knowledge of variation, knowledge of psychology, and knowledge of how knowledge grows" (Berwick, cited in Langley et al., 2009, p. xii). The profound knowledge needed to improve students' mathematical sense-making might include, for example, knowledge about the variability in mathematics instruction in a district and what causes it, knowledge about how to sustain educators' motivation to improve instruction, and knowledge of organizational routines that allow educators to build and share knowledge about instruction. The system of profound knowledge includes both generalizable knowledge (e.g., the impact of intrinsic vs. extrinsic rewards) and organization-specific knowledge (e.g., the incentives within a particular organization to build students' mathematical sense-making).

The Improvement Guide (Langley et al., 2009), at nearly 500 pages, provides one major compendium of improvement science tools and processes, and identifies as the core framework of improvement science the plan-do-study-act (PDSA) cycle, a process for rapid cycles of learning from practice, coupled with three fundamental questions that drive improvement work:

---

[1]Mills College School of Education, Oakland, CA

Educational Researcher, Vol. 44 No. 1, pp. 54–61
DOI: 10.3102/0013189X15570388

1. What are we trying to accomplish?
2. How will we know that a change is an improvement?
3. What change can we make that will result in improvement?

A district interested in improving students' sense-making in mathematics might use tools and processes from *The Improvement Guide* to build buy-in to a shared improvement "charter" (Langley et al., 2009, p. 90) and to build a cause-and-effect diagram (Langley et al., 2009, p. 429) that helps teachers and administrators "see" the system that produces mathematics instruction from each other's viewpoints and identify potential drivers of improvement. Creating such a system diagram might help administrators notice, for example, that the current teacher evaluation system downgrades teachers if students use mathematically imprecise vocabulary, providing a *disincentive* for teachers to elicit students' mathematical explanations. The PDSA cycle provides a way to learn how a change works on a small scale before trying it on a large scale. Other improvement science tools and processes include data display and analytic strategies (such as run charts and control charts) that reveal the extent and causes of variation within a system (for example, whether particular curriculum materials or instructional elements are associated with sense-making) and "practical measurement" strategies that allow evaluation of potential improvement ideas during cycles of rapid prototype development and testing (Yeager et al., 2013). Improvement science can be used at various grain sizes of a system, such as in a single department, a whole organization, or a group of organizations. For simplicity, the term *organization* will be used here.

Why Is Improvement Science of Interest to Educators?

Experimental science, with its hallmark method of the randomized controlled trial (RCT), provides a gold standard for drawing causal inferences and thereby building basic knowledge. But experimental science draws causal conclusions by minimizing variation in both treatment and setting (Lipsey, 1993), for example, by requiring faithful implementation of a program in settings with carefully selected characteristics. Unfortunately, variation is the primary issue that needs to be understood in educational improvement (Bryk, Gomez, & Grunow, 2010). Requiring faithful implementation of a program assumes that the needed knowledge is "in" the intervention and ignores the role of the system of profound knowledge in producing success or failure. Improvement science, in contrast, treats variation in implementation and setting as important sources of information and provides tools to grasp and learn from variation (in both positive and negative directions) in order to redesign both the intervention and the system. As Bryk et al. (2010) note, "rather than thinking about a tool, routine or some other instructional resource as having proven effectiveness, improvement research directs efforts toward understanding how such artifacts can be adaptively integrated with efficacy into varied contexts" (p. 25).

Improvement science and experimental science thus bring different assumptions to scale-up, shown in Figure 1. Experimental science assumes scale-up occurs through faithful implementation of a proven program in new settings. Improvement science assumes scale-up occurs through integration of basic knowledge with the "system of profound knowledge," such as knowledge about how to build shared ownership of improvement, to detect and learn from variations in practice, to build and share knowledge among practitioners, to motivate frontline innovators, and so forth. Some of this knowledge may be embedded in the organization's routines: Many education researchers have noted that organizational and system factors crucially shape program implementation (e.g., Cobb, McClain, de Silva Lamberg, & Dean, 2003; Gutierrez & Penuel, 2014; Spillane, Parise, & Sherer, 2011) and have argued for research on the conditions that allow research-based knowledge to produce improved practice (Coburn & Stein, 2010). Yet there is relatively little education research in the improvement science tradition, which emphasizes building organization members' understanding of the problem and its causes, buy-in to improvement, identification of improvement ideas within and outside their organization, and rapid testing of promising ideas through PDSA cycles.

The ideas underlying improvement science are not new to education researchers. Like improvement science, action research (e.g., Argyris, Putnam, & Smith, 1985) often focuses on identification, analysis, and remediation of a problem in a specified context, often using a process like the PDSA cycle to enact and study change. Like improvement science, the fields of formative (or "theory-driven") evaluation (e.g., Chen, 1990; Donaldson, 2002), design-based research (e.g., Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003) and design-based implementation research (Penuel, Fishman, Haugan Cheng, & Sabelli, 2011) often focus on how basic disciplinary knowledge interacts with organizational conditions to produce improvement. These substantial bodies of work remind us that education researchers have long recognized the interplay of disciplinary knowledge with organizational conditions. Improvement science may contribute new tools and approaches that have proven useful outside education and organizational processes that can be used to build local knowledge in many different settings.

Examples of Improvement Science in Education

Two examples will be used to illustrate the potential of improvement science within education. The first example, the Community College Pathways Networked Improvement Community (NIC; Bryk et al., 2010), developed by the Carnegie Foundation for the Advancement of Teaching, exemplifies improvement science in several ways. It focuses on a shared practical improvement aim—increasing the proportion of students who successfully complete developmental mathematics at member colleges—rather than primarily on advancing theory or testing a predefined program (Silva & White, 2013). The NIC participants (community colleges) use a shared theory of change that identifies a solution *system* (not a single solution) and they measure interim progress by a set of agreed-upon indicators, such as student attendance and attitudes, that measure key elements of their theory of change and can be readily collected by sites. Participants use improvement science tools to "see" the organization and system in which they operate, for example, to notice how many students drop out after

Experimental Science Paradigm **Improvement Science Paradigm**

Example of Knowledge to be Scaled Up: “Proven” Program of Instructional Strategies to Build Students’ Mathematical Sense-Making

	Implement the Program with Fidelity at New Sites	**Integrate the Program with Local Knowledge-Building Systems at New Sites**
Nature of Scale-up	Monitor fidelity of implementation at new sites Design or approve customizations of program if needed Use incentives to achieve implementation if needed	Expect modifications (of the program and site) as the program is integrated with local knowledge-building systems. Key activities might include: ▪ Building consensus on the importance of mathematical sense-making and how to measure its improvement ▪ Mapping the drivers of mathematical sense-making in the current system and how the program would change those; understanding variability in sense-making ▪ Using rapid PDSA cycles to enact and study program elements, refining them as warranted
Assumptions	Knowledge is “in” the program Improvement occurs through faithful program implementation Variation is problematic	Knowledge is also “in” people and systems that use the program Program may need modification, driven by ongoing practical measurement Variation can be source of ideas to improve program and site
Measurement	Use well-validated tools to measure implementation and impact	Use practical measurement tools to test leading indicators predicted by theory of change; use “balancing” measures to check for adverse impacts
Optimal Improvement Conditions	*Sites:* Success most likely if new sites are similar to sites where program was proven *Program/Tool:* Easy to transport and implement, foolproof	*Sites:* New sites need not be similar; success depends on organizational knowledge-building systems as well as program fit *Program/Tool:* Supports knowledge-building, motivation, ownership & customization in many settings

FIGURE 1. *Scale-up of knowledge: Contrast of paradigms*

the first class or the first semester at their institution, compared to other institutions, and to notice strategies used by more successful instructors and colleges (such as buddy activities at the first class or designing a yearlong course rather than separate semester courses). Short PDSA cycles are used to test potential improvements, such as a “group noticing routine” (based in psychological research on belonging) that helps students build interactions outside the immediate mathematics context and is designed to build belonging, mutual sense of responsibility, and attendance among students (Silva & White, 2013; Yeager et al., 2013). The Community College Mathematics Pathways NIC shows remarkable early results, with students earning mathematics college credit at 2 to 3 times the typical rate in roughly half the time (Van Campen et al., 2013).

Lesson Study as Improvement Science

The second example of improvement science, lesson study, occurs in more than 95% of Japanese public schools (National Education Policy Research Institute, 2011). Lesson study in Japan illustrates how improvement science in education can be practiced across a whole nation, scaling up improvement by a route different from that shown for RCTs in Figure 1. The left-hand side of Figure 2 shows the basic lesson study cycle, a collaborative process in which a team of teachers plans, enacts, and examines an intended improvement to instruction (Lewis & Hurd, 2011). Figure 2 also shows four basic types of lesson study that interact synergistically in Japan: lesson study based in schools, districts, university-attached lab schools, and professional associations (Lewis &

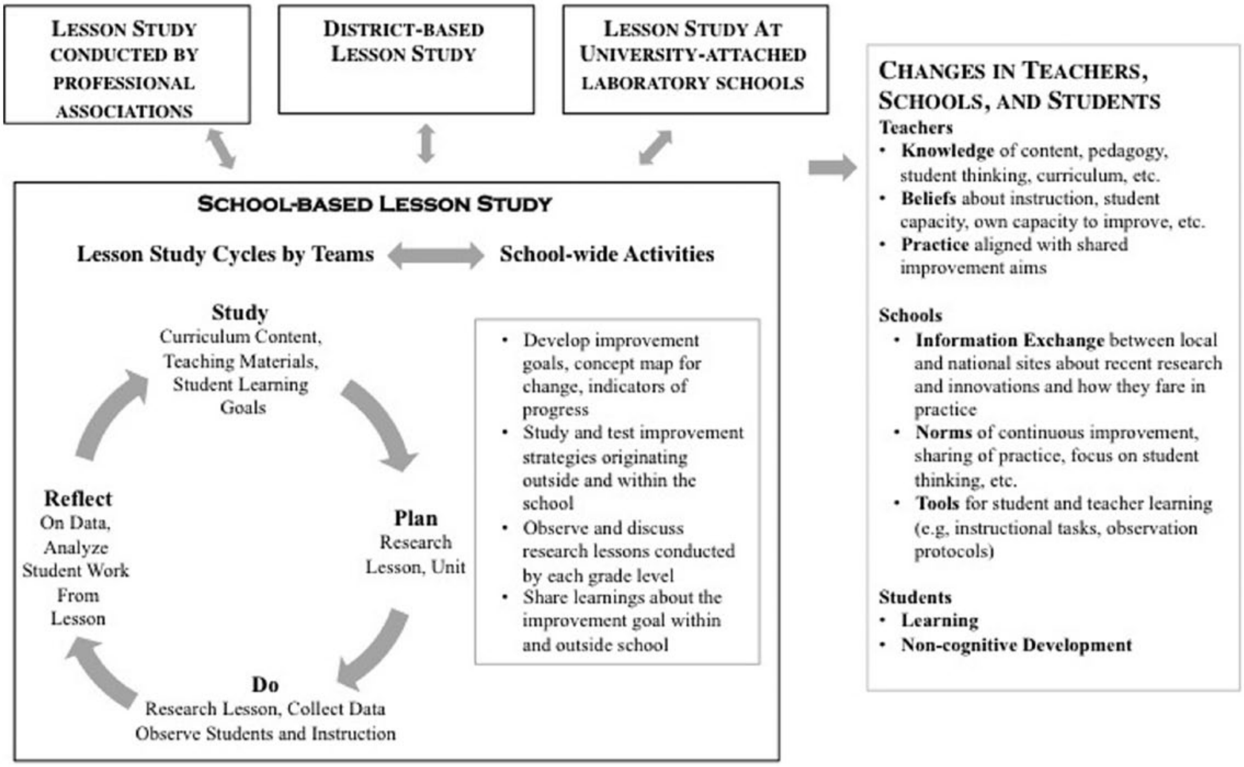

FIGURE 2. *Improvement science at scale: Lesson study in Japan*

Tsuchida, 1997). Together, these different forms of lesson study produce a locally managed, yet nationwide, system in which classroom educators lead the enactment, study, refinement, and spread of instructional improvements.

All four types of Japanese lesson study center on the fundamental improvement science questions listed earlier: Educators choose an improvement aim, agree on how they will recognize improvement, identify the changes that might produce improvement, and test these changes in lesson study cycles (a form of PDSA cycle).

Although the four types of lesson study share the same core questions and lesson study cycle, they focus on different layers of the educational system. The goal of school-based lesson study is to improve instruction at a single school. The other types of lesson study focus on larger systems—the district, region, or even nation as a whole—addressing questions of interest at those levels. For example, lesson study sponsored by a district might focus on improving transition from elementary to junior high school, whereas lesson study sponsored by a national mathematics professional association might investigate whether students are best equipped to understand triangle area or parallelogram area after they study rectangle area—a consequential issue for educators designing curriculum and standards.

A case from Matsuzawa Elementary School in Tokyo (Takahashi & McDougal, 2014) examines how Japanese educators used school-based lesson study to study and enact new national elementary mathematics standards that substantially increased the emphasis on mathematical thinking and explanation. After studying the new standards, Matsuzawa educators chose as their improvement aim "mathematics teaching that helps students explain their ideas to each other and learn from each other." They developed a cause-and-effect diagram (a typical tool of improvement science) to share knowledge with each other about current practices and to map potential improvements to be tested (Takahashi & McDougal, 2014). Their diagram highlighted actions that individual teachers and lesson study groups could take, such as anticipating student thinking, capturing the flow of students' mathematical ideas on the board to support student thinking during lessons, comparing textbooks from several companies to plan a research lesson, and using student journal prompts designed to promote reflection. The Matsuzawa educators also agreed on indicators to gauge the progress of their improvement work. For example, they agreed to use research lessons and student mathematical journals to observe the extent of student explanations and to measure indicators of explanation quality, such as whether students explain in logical steps and explicitly connect diagrams, equations, and words as they explain. A demonstration lesson and lectures by outside mathematics educators helped the Matsuzawa educators build their ideas about instruction that supports student explanations. These outside educators brought in ideas from research and from other lesson study venues.

Grade-level teams at Matsuzawa Elementary School each chose one mathematics unit to study in depth and selected one lesson from the unit as a research lesson to be observed by the whole faculty. The eight resulting research lessons each focused on the schoolwide research theme. Teachers within the school rapidly built and shared knowledge about fostering high-quality student-to-student explanations, for example, knowledge about effective journal prompts and teacher questions.

School-based lesson study, like that at Matsuzawa School, does not exist in isolation. It influences and is influenced by lesson study conducted in other settings. For example, all Matsuzawa teachers also participate in district-based lesson study, choosing among a dozen or so subject offerings and working in a cross-school lesson study group with other district teachers particularly interested in that subject. Local and national professional organizations and university lab schools also conduct lesson study; their lesson study participants typically have good access to recent worldwide research and strong expertise in a particular discipline as well as positions in schools where they can enact these ideas (Lewis & Tsuchida, 1997; Murata & Takahashi, 2002; Shimizu, 2002b). Ideas flow among the different lesson study venues in many ways. Annual public research lessons sponsored by university lab schools and by professional associations may attract a thousand or more educators; research lessons are amplified and projected for these large audiences. Well-known educators, like those who provided a demonstration lesson and lectures at Matsuzawa School, also spread ideas from site to site, serving as "boundary crossers" (Akkerman & Bakker, 2011) as they comment on lessons sponsored by university lab schools, professional associations, and local schools. Likewise, live research lessons and publications (lesson plans, lesson reports, and video) spread information and serve as "boundary objects" that allow groups to negotiate a shared vision of approaches, like "teaching through problem solving" (Shimizu, 2002a; Takahashi, 2008). Both university-based and school-based participants may be changed by participation in these boundary events: Local lesson study groups gain access to recent innovations and feedback on their own use of them in research lessons, and university-based educators can learn how research-based innovations are taken up (or not) across varied settings and can see both successful refinements and lethal mutations.

Through the synergy of ongoing lesson study in local schools, conducted by teachers who know their local students and setting well, and lesson study at regional and national levels, conducted by educators who have access to recent research and also to school settings where they can enact it in classrooms, Japanese education has made broad, teacher-led shifts from "teaching as telling" to teaching for understanding in both mathematics and science, successfully spreading some major instructional innovations that were developed, but never spread widely, in the United States (Lewis & Tsuchida, 1997; Stigler & Hiebert, 1999; Takahashi, 2008).

A Future for Improvement Science Within Educational Research?

The two cases illustrate how educators can improve instruction by integrating basic disciplinary knowledge (e.g., for example, research on belonging or on students' mathematical explanations) with organizational processes, such as development of a shared improvement aim, cause-and-effect mapping to share current practice and identify potential drivers of improvement, and PDSA cycles to test potential improvements. Both improvement science (Park, Hironaka, Carver, & Nordstrum, 2013) and lesson study (Akiba, Ramp, & Wilkinson, 2014; Hill, 2011) have spread rapidly in the United States, showing promising results on measures like attendance and course completion (Park et al., 2013; Van Campen, Sowers, & Strother, 2013; Yeager et al., 2013). A recent review of 643 studies of mathematics professional development, using a process modeled on What Works Clearinghouse guidelines, found only two studies that met scientific criteria and showed impact on student learning, and one of them was lesson study supported by mathematical resources (Gersten, Taylor, Keys, Rolfhus, & Newman-Gonchar, 2014; Lewis & Perry, 2014). Recent reconceptualizations of education research argue the importance of building educators' capacity to adapt research-based ideas (e.g., Gutierrez & Penuel, 2014) and to learn from variation (Bryk et al., 2010), goals for which improvement science is well suited. Four changes within education research might enhance our capacity to benefit from improvement science.

1. Recognize the different affordances of experimental and improvement science. Looking at Figure 1, it is easy to imagine educational programs suited to research under each paradigm. When a proposed improvement does not interact with system knowledge—for example, when teachers can individually implement a change using a packaged program, without coming into conflict with system features (such as assessment or curriculum requirements)—an RCT allows causal inferences about a reproducible intervention. However, a goal such as building students' sense-making in mathematics is likely to interact substantially with profound knowledge—to require adjustment of system elements, such as curriculum, evaluation systems, and assessment, along with attention to sustaining teachers' motivation for the hard work of changing practices and beliefs. The way we currently talk about scale-up often seems to assume that RCTs always offer the best design for building generalizable knowledge, but this assumption may hold only for highly specified programs that do not interact heavily with profound knowledge. When variation in setting is likely to affect implementation, or when will and capacity need to be built, the RCT may offer an overly simplistic research design, rooted in the faulty assumption that basic disciplinary knowledge is sufficient to produce improvement. Instead, we may need to focus on building additional types of generalizable knowledge, as the next section explores.

2. Recognize different types of generalizable knowledge. Knowledge for improvement may be captured in tools, "actionable artifacts" (Bannan-Ritland, 2003), and descriptions of change processes; however, published education research often shortchanges such process-related information in favor of results (Goldsmith, Doerr, & Lewis, 2014; Sztajn, 2011). Knowledge for improvement may also be embodied in organizations and individuals, for example, in organizations that know how to learn from variations in practice and conduct rapid, small tests of promising ideas (Bryk et al., 2010) and in educators who notice and use student thinking (Sherin, Jacobs, & Philipp, 2011). Lesson study in Japan exemplifies a set of improvement processes that have generalized across many different settings and purposes. Special journal issues and conferences could draw attention to these forms of generalizable improvement knowledge and how they are best spread, for example, how organizations learn to conduct PDSA cycles, nurture careful attention to student thinking, and build will for shared improvement.

3. Create a place for practical measurement. PDSA cycles are often driven by "practical measurement" that can feed back information quickly and without disrupting ongoing activity of the system (Yeager et al., 2013). A small number of items may be used, some items may target organization-specific processes, scale reliability may be unknown, and causal links with valued outcomes may not yet be established. Nevertheless, the items target expected change processes and may be better suited to improvement research than are established, validated measures, since practical measures are brief, are specific to the change theory, and often include "balancing" measures that check on potential negative consequences of innovation (Langley et al., 2009). For example, in addition to using students' journals and classroom speech during research lessons to gauge improvement in student explanations, the Matsuzawa teachers also measured students' enjoyment of explaining their ideas, as a "balancing" measure to detect any adverse effects on students' enjoyment of mathematics.

4. Learn across boundaries. Improvement science is striking for the dramatic examples of cross-industry, cross-national learning that have occurred (Berwick, 2003; Gawande, 2007; Rother, 2009). Why have improvement scientists been able to transfer innovations across contexts as different as automobile manufacturing and medicine, as Asia and North America, when it is typically so hard to transfer educational innovations across school and district boundaries in the United States? As Figure 1 highlights, an assumption of the experimental science paradigm is that scale-up requires careful specification and faithful implementation of an intervention. In contrast, improvement science assumes that the knowledge contained in the intervention will need to be activated by knowledge-building systems within an organization, for example, strategies to mobilize a sense of shared purpose and routines to detect and learn from variation. In experimental science thinking about scale-up, "lethal mutations" of an intervention (Brown & Campione, 1994) are avoided by thoroughly specifying the intervention and monitoring implementation fidelity. In improvement science thinking, lethal mutations are avoided (in the long term) by building capacity within the organization to understand the factors that shape improvement, to notice and learn from variation, and to build frontline implementers' ownership of change (Bryk et al., 2010; Langley et al., 2009). Once ownership and capacity to learn from variation and to improve practice develop *within* an organization, organization members can learn from programs and ideas rooted in quite different contexts; in fact, different cultures or organizational sectors may provide the best opportunity to identify assumptions that go unrecognized within one's own setting. Improvement science thus opens up the possibility of using educational practices from very different settings.

U.S. teachers often respond enthusiastically to study of Japanese textbooks, lesson videos, and the lesson study process itself, as captured in the following reflections written by individual teachers at the end of lesson study meetings:

> Looking at the Japanese books was really eye opening. They were so well laid out, working with several ways to attack a problem; [they] really seemed to take time to develop number sense where we push formulas. It is such an exciting way to look at math. As a group we discussed how the Japanese teach comfort with making mistakes and risk taking. . . . The idea of not giving formulas until meaning has been constructed by individuals is different from our "normal" lesson presentation. (ID368)

> The whole process of lesson studies is much clearer. . . . I learned that we aren't creating something new, but rather looking for & taking what is already out there and making it better! I am much more aware of how little time is devoted to collaborative work & observation of other teachers in the U.S. SAD! . . . We spend way too much time on the islands we call our classrooms. . . . We constantly "recreate the wheel" rather than drawing upon the skills & expertise of fellow teachers! I like that every team member has ownership in the process. (ID370)

In contrast to the enthusiasm of these teachers and the evidence of significant impact of lesson study with Japanese mathematical resources on U.S. teachers' and students' mathematical knowledge (Gersten et al., 2014; Lewis & Perry, 2014), education researchers often express skepticism about cross-cultural borrowing. For example, one journal reviewer cited provision of materials from the Japanese teacher's edition to U.S. lesson study groups as evidence of "uncritical adoption of a Japanese classroom practice . . . Pedagogical methods are culturally embedded and . . . transplanting them from one culture to another is not always feasible" (Hatano & Inagaki, 1998, pp. 101–102). These arguments may be important in an experimental science paradigm in which educators are required to implement a program faithfully but less important under the improvement science assumption that organizations must integrate a program with their own knowledge-building systems. Arguing that reforms from around the world provided building blocks for Finland's educational improvement, Andy Hargreaves notes,

> We have spent decades breaking down the isolation of teachers within and between our schools. It is now time to break down the ideology of exceptionalism in the United States and other Anglo-American nations if we are to develop reforms that will truly inspire our teachers (cited in Sahlberg, 2011, p. xx).

Experimental science has produced much important educational knowledge that can be spread and refined by improvement science methods, in much the way that Japanese educators spread innovations (some from the United States) across Japan, through locally managed improvement cycles (Lewis & Tsuchida, 1997; Murata & Takahashi, 2002; Takahashi, Lewis, & Perry, 2013). Education research would benefit from inclusion of improvement science, which has methods tailored to rapid prototyping and testing, tools for detecting and learning from variation, and affordances to learn from widely different contexts. Improvement science recognizes the capacity of practitioners to engage in disciplined inquiry using tools and ideas from another site and not only to faithfully implement a researcher-designed program. Improvement science also recognizes the right of practitioners to use currently available innovations, rather than waiting for a "proven" program. Finally, improvement science recognizes that educational improvement will not occur solely through advances in basic disciplinary knowledge but will also require advances in the "the system of profound knowledge" needed to enact it.

NOTES

This work was supported by the Institute for Education Sciences, U.S. Department of Education, under Grant Nos. R308A960003, R305A070237, and R305A110491; the National Science Foundation under Grant No. 0207259; and the Toyota Foundation. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the Institute for Education Sciences, the National Science Foundation, or the Toyota Foundation. I wish to thank Jim Stigler, Tony Bryk, Corey Donahue, and anonymous reviewers for their suggestions on this essay.

[1] I use *improvement science* because this phrase (or the longer phrase *the science of improvement*) is commonly used by its practitioners, for example, Langley et al. (2009), who note the dictionary definition of *science* as "knowledge attained through study or practice."

REFERENCES

Akiba, M., Ramp, L., & Wilkinson, B. (2014). *Lesson study policy and practice in Florida: Findings from a statewide district survey*. Tallahassee: Florida State University.

Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research*, *81*(2), 132–169.

Argyris, C., Putnam, R., & Smith, D. (1985). *Action science: Concepts, methods and skills for research and intervention*. San Francisco, CA: Jossey-Bass.

Bannan-Ritland, B. (2003). The role of design in research: The integrative learning design framework. *Educational Researcher*, *32*(1), 21–24.

Berwick, D. (2003). Improvement, trust, and the healthcare workforce. *Quality and Safety in Health Care*, *12*(6), 448–452.

Brown, A. L., & Campione, J. C. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 229–270). Cambridge, MA: MIT Press.

Bryk, A. S., Gomez, L. M., & Grunow, A. (2010). *Getting ideas into action: building networked improvement communiites in education*. Retrieved from http://www.carnegiefoundation.org/spotlight/webinar-bryk-gomez-building-netowrked-imporvement-communities-in-education

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.

Chen, H. T. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*(1), 9–13.

Cobb, P., McClain, K., de Silva Lamberg, T., & Dean, C. (2003). Situating teachers' instructional practices in the institutional setting of the school and district. *Educational Researcher*, *32*(6), 13–24.

Coburn, C. E., & Stein, M. K. (2010). *Research and practice in education: Building alliances, bridging the divide*. Lanham, MD: Rowman & Littlefield.

Common Core State Standards Initiative. (2010). *The standards: Mathematics*. Retrieved from http://www.corestandards.org/the-standards/mathematics

Donaldson, S. (2002). Theory-driven program evaluation in the new millennium. In S. I. Donaldson & M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium* (pp. 115–146). Mahwah, NJ: Lawrence Erlbaum.

Gawande, A. (2007). *Better: A surgeon's notes on performance*. New York, NY: Picador.

Gersten, R., Taylor, M. J., Keys, T. D., Rolfhus, E., & Newman-Gonchar, R. (2014). *Summary of research on the effectiveness of math professional development approaches*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Educational Evaluation and Regional Assistance, Regional Educational Laboratory Southeast.

Goldsmith, L. T., Doerr, H. M., & Lewis, C. C. (2014). Mathematics teachers' learning: A conceptual framework and synthesis of research. *Journal of Mathematics Teacher Education*, *17*(1), 5–36. doi:10.1007/s10857–013–9245–4

Gutierrez, K. D., & Penuel, W. R. (2014). Relevance to practice as a criterion for rigor. *Educational Researcher*, *43*(1), 19–23.

Hatano, G., & Inagaki, K. (1998). Cultural contexts of schooling revisited: A review of the learning gap from a cultural psychology perspective. In S. G. Paris & H. M. Wellman (Eds.), *Global prospects for education development, culture and schooling* (pp. 79–104). Washington, D.C: American Psychological Association.

Hill, H. C. (2011). The nature and effects of middle school mathematics teacher learning experiences. *Teachers College Record*, *113*(1), 205–234.

Langley, G. J., Moen, R. D., Nolan, K. M., Nolan, T. W., Norman, C. L., & Provost, L. P. (2009). *The improvement guide*. San Francisco, CA: Jossey-Bass.

Lewis, C., & Hurd, J. (2011). *Lesson study step by step: How teacher learning communities improve instruction*. Portsmouth, NH: Heinneman.

Lewis, C., & Perry, R. (2014). Lesson study with mathematical resources: A sustainable model for locally-led teacher professional learning. *Mathematics Teacher Education and Development*, *16*(1).

Lewis, C., & Tsuchida, I. (1997). Planned educational change in Japan: The case of elementary science instruction. *Journal of Educational Policy*, *12*(5), 313–331.

Lipsey, M. W. (1993). Theory as method: Small theories of treatments. *New Directions for Program Evaluations*, *57*, 5–38.

Murata, A., & Takahashi, A. (2002). Vehicle to connect theory, research and practice: How teacher thinking changes in district-level lesson study in Japan. In D. L. Haury (Ed.), *Proceedings of the Twenty-Fourth Annual Meeting of North American Chapter of the International Group of the Psychology of Mathematics Education* (pp. 1879–1888). Columbus,OH: ERIC Clearninghouse for Science, Mathematics, and Environmental Education.

National Education Policy Research Institute. (2011). *Report of survey research on improvement of teacher quality* [Kyouin no Shitsu no Koujou ni Kansuru Chosa Kenkyuu]. Tokyo, Japan: National Education Policy Research Institute.

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

Penuel, W. R., Fishman, B. J., Haugan Cheng, B., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher*, *40*, 331–337.

Rother, M. (2009). *Toyota Kata: Managing people for improvement, adaptiveness and superior results*. New York, NY: McGraw-Hill Professional.

Sahlberg, P. (2011). *Finnish lessons: What can the world learn from educational change in Finland?* New York, NY: Teacher's College Press.

Sherin, M., Jacobs, V., & Philipp, R. (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. London, UK: Taylor & Francis.

Shimizu, Y. (2002). Capturing the structure of Japanese mathematics lessons: Some findings of the international comparative studies. In D. Edge & Y. B. Har (Eds.), *Proceedings of Second East Asia Regional Conference on Mathematics Education and Ninth Southeast Asian Conference on Mathematics Education: Vol. 1. Invited papers:*

Plenary lectures and regular lectures (pp. 168–176). Singapore: National Institute of Education.

Shimizu, Y. (2002, July). *Sharing a new approach to teaching mathematics with the teachers from outside the school: The role of lesson study at "Fuzoku" schools*. Paper presented at the US-Japan Cross Cultural Seminar on the Professionalization of Teachers Through Lesson Study, Park City, UT.

Silva, E., & White, T. (2013). *Pathways to improvement: Using psychological strategies to help college students master developmental math*. Retrieved from http://www.carnegiefoundation.org/resources/publications/pathways-improvement-using-psychological-strategies-help-college-students-master-developmental-math/

Spillane, J. P., Parise, L. M., & Sherer, J. Z. (2011). Organizational routines as coupling mechanisms: Policy, school and administration, and the technical core. *American Educational Research Journal*, *48*(3), 586–620.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York, NY: Summit Books.

Sztajn, P. (2011). *Standards for reporting mathematics professional development in research studies*. Reston, VA: National Council of Teachers of Mathematics.

Takahashi, A. (2008, July). *Beyond show and tell: Neriage for teaching through problem-solving. Ideas from Japanese problem-solving approaches for teaching mathematics*. Paper presented at the 11th International Congress on Mathematical Education, Monterrey, Mexico.

Takahashi, A., Lewis, C., & Perry, R. (2013). Facilitating curriculum reforms through lesson study. *International Journal for Lesson and Learning Studies*, *2*(3), 207–217.

Takahashi, A., & McDougal, T. (2014). Implementing a new national curriculum: Case study of a Japanese school's 2-year lesson-study project. In K. Karp (Ed.), *Annual perspectives in mathematics education: Using research to improve instruction 2014* (pp. 13–22). Reston, VA: National Council of Teachers of Mathematics.

Van Campen, J., Sowers, N., & Strother, S. (2013). *Community College Pathways: 2012–13 descriptive report*. Stanford, CA: Carnegie Foundation for the Advancement of Teaching.

Yeager, D., Bryk, A., Muhich, J., Hausman, H., & Morales, L. (2013). *Improvement research carried out through networked communities: Accelerating learning about practices that support more productive student mindsets*. Paper presented at the White House Meeting on Excellence in Education: The Importance of Academic Mindsets, Washington, DC. Retrieved from http://cdn.carnegiefoundation.org/wp-content/uploads/2014/09/improvement_research_NICs_bryk-yeager.pdf

AUTHOR

CATHERINE LEWIS, PhD, is a distinguished research scholar at Mills College School of Education, 5000 MacArthur Boulevard, Oakland, CA 94613; *clewis@mills.edu*. Her research (www.lessonresearch.net) focuses on professional learning and on Japanese elementary education.

Manuscript received April 11, 2014
Revisions received November 1, 2014,
and January 6, 2015
Accepted January 8, 2015