The Legitimization of Improvement Science in Academe

Casey D. Cobb[1] and Patricia Virella[2]

[1,2]*University of Connecticut*

[1] *casey.cobb@uconn.edu*

Abstract

This article examines improvement science (Bryk, 2009) against the backdrop of traditional academic research methods. Improvement science is perhaps most closely aligned with design-based implementation research, and is typically applied to networked communities (e.g., schools, hospitals) with the goal of continuous organizational improvement. Improvement science has earned value among the practitioners and researchers who engage in it, but still seeks more complete legitimacy within the academy. We describe the method of improvement science and situate it within the two paradigms of design research and research design. Examples of its implementation in school reform and university program improvement are shared to illuminate the systematic and dynamic nature of its process. The article speaks to the normalization of research design and design research within the context of "what counts" as research in academia, and where and how improvement science can fit within these traditions.

Keywords*: improvement science, research design, design-based research, school reform, education*

1. Introduction

In this article, we explore the disconnect between research and practice in the field of education from the perspective of the methods largely employed. Although traditional forms of research design remain dominant, they are increasingly giving ground to new models of *design research*. We discuss an offshoot of design research, called improvement science, and illustrate its features and applications. Like design research, improvement science offers an alternative to the separation between researcher and practitioner that contributes to the scholarship-practice divide. Although the examples we use here are in the context of educational organizations, improvement science is used by other disciplines, including business, healthcare, architecture, and engineering. Thus readers from other fields may see its potential implementation in their own settings. The article concludes with a discussion on the challenges for improvement science to gain legitimacy in academe.

2. The Research-Practice Divide

In our discipline of education, there continues to be a persistent gap between research and practice (Broekkamp & van Hout-Wolters, 2007; Dynarski, 2015;

Nuthall, 2004). Other fields in the social sciences also face this challenge, but it seems acutely the case in education. Research in areas such as student learning, school reform, and teacher effectiveness tends to yield knowledge incrementally, and new understandings are not necessarily translated into practice.

Many factors impede the translation of research into practice and maintain distance between scholar and practitioner. From the perspective of practitioners, it may be due to the inaccessibility of complex analytic methods, the lack of confidence that findings will implement in their settings, or the notion that researchers are out of touch with the current realities of schools. From the standpoint of researchers, it could be because prevailing research paradigms call for the researcher to remain detached, objective, and unbiased. Furthermore, university promotion and tenure systems reward faculty for publications in high-impact journals. Often times these are not even read by practitioners who are engaged in the real work of schooling.

The vast majority of scholarship in education follows conventional research designs undergirded by qualitative, quantitative, or mixed method paradigms. For the most part, these traditional forms of scholarship position the researcher as knowledge generator and the practitioner as knowledge consumer. Even systematic research designs that are carefully executed have substantial limitations. Consider the classic experimental design applied in a school setting. For example, a researcher may be interested in assessing the effects of a new literacy program on the reading performance of middle school students. A basic experiment would call for the random assignment of students to treatment and control groups—a design rarely possible given the intact nature of classrooms and the disruption that would invariably ensue in displacing students. A viable alternative is to randomize at the classroom level. But even here researchers face numerous threats to internal validity that would be difficult to overcome. These include effectively controlling for teacher and student factors that could bear upon the outcome measures (e.g., teacher skill, ability to implement the program with fidelity). Good researchers would attempt to address threats through manual or statistical matching and other statistical controls, but even these techniques are not without their own limitations. In the end, the experiment would yield findings appropriately qualified by "study limitations." In reality, even using the experimental design, it is extremely difficult to attribute observed effects solely to an educational intervention. Education settings are complex social environments. Consider the multitude of internal and external influences on student performance that fall outside the specific intervention. Moreover, what is typically reported from such studies are *average* treatment effects. Averages do not take into account differential treatment effects on students based on factors such as student learning styles, reading abilities, and internal motivation. Finally, beyond the difficulty in overcoming threats to internal validity are limitations to the study's external validity. External validity speaks to the generalizability of a study's findings beyond the participants and settings in which the study took place. Experimental research may be the "gold standard" of empirical designs seeking to confer causality (Ginsburg & Smith, 2016), but education is not configured to take advantage of its power.

To be sure, there are several other systematic, rigorous research designs intended to inform practice. These include other more targeted forms of experimental research, such as repeated measures and single subject designs, and sophisticated multivariate modeling, such as hierarchical linear regression.

Interpretivist approaches found in qualitative designs aim to generate rich understanding of how students learn and under what conditions. These methods include ethnography, grounded theory, and phenomenology. These two approaches bring the researcher closer to the naturalistic settings and realities of schools. But these designs are still mainly aimed at producing generalizable or transferable findings to outside entities and settings. They are not necessarily aimed at—or sufficiently equipped—to improve practice in specific settings in the here and now.

Many of the major grant agencies and foundations also place high value on the production of traditional forms of scholarship and produced through the application of rigorous research designs. For the past couple of decades, the U.S. Department of Education's Institute for Education Sciences (IES) has placed a premium on experimental designs in its various grant programs. IES is the custodian of the *What Works Clearinghouse*, which only lists educational interventions that have been proven effective by rigorous causal research designs, such as randomized controlled trials, regression discontinuity, and other approved quasi-experimental models.[1]

In response to the research-to-practice disconnect in our field, other paradigms of inquiry have emerged, such as design-based research, action research, participatory research, implementation research, and utility-focused evaluation. All are attempts to bring research closer to the action with the goal of ensuring research is usable. An approach that holds particular promise is design-based research. We turn next to describe design-based research and contrast it with traditional research design.

3. Design-based Research and Traditional Research Design

Design research (Oha & Reeves, 2010), or what it has become more popularly known as design-based research (McKenney & Reeves, 2013), is a relatively new approach in education research. Design-based research came on in full force at the turn of this century with promise to better bridge research and practice (Design-Based Research Collective, 2003). Between 2003 and 2004, at least four prominent education journals devoted themed issues to illustrate its features, purposes, and applications (Anderson & Shattuck, 2012, Tiberghien, Vince, & Gaidioz, 2009).

Design-based research is quite unlike traditional research design. Conventional quantitative research designs adhere to positivist and post-positivist epistemologies; they typically identify a topic needing study, review the extant research and theory, pose research questions and hypotheses, and systematically collect and analyze data to test those hypotheses. Research designs undergirded by an interpretivist epistemology follow a similar strategy, but they are less interested in testing hypotheses and more disposed to learning inductively from data collected among participants and across settings. In contrast, design-based research is intended to be pragmatic and focused on improvement. Design-based research is considered

[1] See the *WWC Version 4.0 Standards Handbook* at https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

> a systematic but flexible methodology aimed to improve educational practices through iterative analysis, design, development, and implementation, based on collaboration among researchers and practitioners in real-world settings, and leading to contextually-sensitive design principles and theories. (Wang & Hannafin, 2005, p. 6)

It is defined by several key characteristics, which include research that (a) is situated in real educational contexts, (b) is focused on the design and assessment of a specific intervention, (c) is open to multiple modes of data collection and analysis, (d) is conducted in iterative stages, (e) involves partnerships between researchers and practitioners, and (f) fosters the evolution of design principles (Anderson & Shattuck, 2012).

Although it is difficult to determine when the "end" of the design-based process has been reached, one marker is whether the inquiry has advanced new theoretical understandings and practical design principles. Design-based research does not attempt to "decontextualize these principles or generate grand theories that function with equal effect in all contexts" (Anderson & Shattuck, 2012, p. 17). In contrast to the utility of grand theories, theories produced from design-based research represent grounded theories that "do real work" (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003, p. 10). The emphasis is on the practical application of new design principles in specific contexts.

4. Improvement Science: An Offshoot of Design-based Research

Improvement science established its roots in healthcare as a means to close what was deemed a "quality chasm" (Junghans, 2018, p. 124). In an effort to make implementation consistent within healthcare, institutions adapted the improvement science framework to make rapid adjustments (Rohanna, 2017). In education, this approach has been heavily promoted by Anthony Bryk and the Carnegie Foundation for the Advancement of Teaching (Bryk, 2009). It has provided educators with the tools to make immediate adjustments to their practice without having to wait for a new school year or directives from central office. Because improvement science shifts the locus of control to practitioners in the field, it allows for instant feedback loops (Bryk, 2015).

Improvement science asserts the value of consistently evaluating, planning, and enacting as a cycle of improvement. Lewis (2015) described the core framework of improvement science as "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. 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?" (p. 54)

These core principles provide organizations with a framework to pivot or maintain course within their organizational design. Moreover, by providing this framework to organizations, replication of this improvement provides an adaptable structure.

Improvement science does not follow a fixed script; rather, it offers a guiding framework and set of inquiry-based activities to advance continuous improvement. Root cause analysis using fishbone charts, driver diagrams to capture theories of action, and plan-do-study-act (PDSA) protocols to conduct rapid cycles of inquiry are common among the main procedures. The most instrumental are the iterative cycles of inquiry. They are intended to learn "what works, for whom, and under what set of conditions" (Bryk, Gomez, Grunow, & LeMahieu, 2015a, p. 2). In its purest sense, improvement science accelerates learning through doing.

Improvement science is an analytic design that can take place at any level of organization but is best done within and across a network of collaborators. Bryk, Gomez, Grunow, and LeMahieu, (2015a) referred to these as "networked improvement communities" or NICs. NICs are similar to professional learning communities (Stoll, Bolam, McMahon, Wallace, & Thomas, 2006) or communities of practice (Wenger & Snyder, 2000), which have become popular approaches to professional development in education. One major difference, however, is that NICs target a narrowly defined problem of practice shared across the network. The power of the network is its ability to accelerate learning and do it at scale.

Networks can lead to what Bryk referred to as "collective knowledge building" (Bryk, Gomez, Grunow, & LeMahieu, 2015a, p. 3). One striking example of this was a World Health Organization NIC composed of doctors, researchers, and medical labs, swiftly assembled to reverse the outbreak of Severe Acute Respiratory Syndrome (SARS) (Bryk, Gomez, Grunow, & LeMahieu, 2015b). The NIC generated solutions to the world epidemic quickly and efficiently through collaborative problem solving and shared learning.

The improvement science framework is a close cousin of design-based research. Their similarities include an emphasis on solving a specific problem in context, repeated testing and retesting of interventions, deployment of multiple methods, and a partnership between researchers and practitioners. Like design-based research, improvement science is concerned with identifying design principles that work in a particular local context.

A networked improvement community engaging in improvement science pursues a vexing problem—a grand challenge—common to the network. NIC collaborators operate, at least initially, with a shared working theory of why the problem exists and test their assumptions with rapid cycles of inquiry. NIC members are encouraged to conduct informal yet organized mini-experiments. Their objective is to understand how and why specific interventions work in specific contexts. The strength of the network is the built-in structures to share results, share them quickly, and share them often. Through its use of NICs, improvement science serves to generate design principles at scale.

Improvement science differs from design-based research in at least one fundamental respect, however. Improvement science relies heavily on shared knowledge and learning within a *community*—a networked improvement community. Design-based research is more apt to isolate the research site and assert a higher degree of scientific control. There is interest in maintaining containment across researchers, practitioners, and participants.

Improvement science also may diverge from the scientific rigor of design-based research. This is not a weakness but rather the reality of conducting real-time, rapid cycles of inquiry across a range of participants, scenarios, and settings. Improvement science relies on inductive and deductive reasoning, to be sure, but also benefits from abductive reasoning. According to Kolko (2010), "abduction can be thought of as the argument to the best explanation. It is the hypothesis that makes the most sense given observed phenomenon or data and based on prior experience. Abduction is a logical way of considering inference or 'best guess' leaps" (p. 20). Consider how similar this approach is to Einstein's infamous "thought experiments" (Brown, 2011).

5. Illustrations of Improvement Science in Education

Lewis (2015) offered two key illustrations that exemplify improvement science in education: The Community College Pathways Networked Improvement Community (NIC) and Matsuzawa Elementary School in Tokyo. In both examples, schools were looking to improve key elements influencing their school environments. The Community College Pathways NIC was trying to improve aspects of the collegiate experiences for their students and Matsuzawa Elementary School used a system of lesson study as a tool to improve their adoption of new mathematics standards. Lewis (2015) posited that both of these cases illustrate how stakeholders use a system already in place to meet the needs of a research query in mind, "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." (p. 58)

The Community College Mathematics Pathways Networked Improvement Community (NIC) used "improvement science tools to 'see' the organization and system in which they operate[d]" (Lewis, 2015, p. 55). The NIC emerged to address a vexing issue in the community college world, where a woeful proportion of students successfully completed developmental math courses. Rather than evaluating the effects of a new curriculum, the NIC rallied around a shared improvement objective. They used 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" (p. 55).

One initial goal of the NIC was to keep students in these introductory math course for the entire first semester. This organization explored these issues by conducting short Plan-Do-Study-Act cycles tethered to protocols such as "group noticing routine" (Lewis, 2015, p. 56). By bringing these elements together, the improvement science protocol helped the organization determine in what ways they could make improvements. Community College Mathematics Pathways NIC determined that they would have year-long courses as a recommendation that resulted from their research inquiry. In this case, the use of improvement science provided short cycles of research and action which allowed the college to make a concrete improvement. "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)" (Lewis, 2015, p. 56).

In Japan, lesson study is used as a mechanism for improvement science. Teachers come together and review their lessons, diving deeply into how to present content to students and reflect upon successes and next steps. Although this may not follow the prototypical PDSA, the elements of improvement science are present within a lesson study. In the case of Matsuzawa Elementary School in Tokyo, the teachers used lesson study as a method to meet the needs of new mathematics standards, which increased the rigor and mathematical explanation students needed to demonstrate in order to meet performance standards. Teachers conducted lessons with a specific aim in mind: that is, "mathematics teaching that helps students explain their ideas to each other and learn from each other" (Lewis, 2015, p.57). The specificity of this aim drove the inquiry and improvement cycle which would guide their lesson study. Educators of this school then agreed upon indicators and criteria for success as well as unpacked units of study within grade levels. Educators within each grade level took ownership of an element of the study.

The results of this lesson study yielded new ways to engage students via journal prompts. Engaging in lesson study as improvement science, educators were able to glean concrete methods of instruction, which improved academic achievement. Moreover, within this example, lesson study becomes the central instrument of improvement science. This demonstrates that improvement science is not simply a framework to be replicated but the foundation of research inquiry within education.

Another illustration of improvement science in action is the long-term partnership between Chicago Public Schools and researchers from University of Illinois at Chicago. Plagued by failing scores and other dismal conditions, Chicago public schools were in need of reform (Bryk, Sebring, Allensworth, Easton, & Luppescu, 2010; Bryk, 2010; Bryk et al., 2015b). Over a seven-year period, Bryk alongside other researchers interrogated the ways in which schools embarked upon the improvement of the district. The framework was adopted at scale in large school districts such as Chicago in an effort to foster change in an underperforming school district. This ensuing case study offered powerful evidence of the effect of improvement science in education. One implication of improvement science within a school's continuous improvement is that it allows a replicable framework that positions several stakeholders as researchers.

The final example of improvement science we share is one being applied across a network of educational leadership scholars. The University Council for Educational Administration (UCEA) represents a consortium of over 100 university programs that prepare educational leaders. In the fall of 2016, UCEA launched its Program Design Network.[2] The Network constitutes five separate NICs, each aimed at a particular problem of practice salient in school principal preparation. These include program design NICs organized around topics such as district-university partnerships, mentorship, coaching, and candidate recruitment.

One of us serves as lead facilitator of the Candidate Recruitment, Selection, and Assessment NIC, a cross-institutional team of faculty from five universities. In our initial work together, we identified recruitment of candidates from historically underrepresented groups as a primary challenge. To tackle the problem, we engaged in root cause analysis, reviewed the extant literature, and shared common challenges in our respective settings. This work prompted several new strategies to try to attract more racially diverse applicants to our pools. The NIC meets regularly throughout the year in both face to face and virtual formats. The process is ongoing and has yielded new insights from which we have all benefited.

Our NIC members are fully committed to increasing the number of underrepresented principal candidates. They do so for the good of their programs and for the good of our overall field. This is part of our jobs as program faculty, so the time is well spent and appropriately accounted for by our university

[2] See http://www.ucea.org/initiatives/program-design-network-pdn/

employers. Our experience as a NIC composed of tenure track faculty has raised questions about how our work can be parlayed into scholarship that is valued by our institutions. It also prompted us to think of other scenarios where university researchers partner with practitioners in the practice of improvement science: how would these efforts and contributions be valued for tenure track faculty?

6. Seeking Legitimacy in Academe

Traditional research design still dominates the landscape of educational research. A premium is placed on creating new knowledge that advances theoretical understanding. Such research follows a common protocol: identifying a problem or gap in the research, providing relevant empirical and theoretical background on the topic, collecting data on particular samples and settings, analyzing data, and discussing the implications of those results for practice and theory.

Although not nearly as prominent as the entrenched traditional research design, design-based research has found its place in educational research (Anderson & Shattuck, 2012). Improvement science, arguably a derivative of design-based research, on the other hand, has not. One reason could be that design-based research still places the trained researcher at the front. Collaboration among researchers, implementers, and practitioners is a fundamental feature of design-based research, but researchers are positioned as the lead. As leaders of the research, they are better able to execute and record research activities that lend themselves to academic publication. In contrast, improvement science is placed primarily in the hands of practitioners; in the context of researcher-practitioner partnerships, there is an equality to the partnership. This is justified in that improvement science is intended to directly improve practice and not as interested in documenting the process post hoc through traditional academic outlets. The focus is on practical learning within a network improvement community to solve common problems of practice.

Improvement science generates specific, practice-based knowledge through what Phelps, Heidl, and Wadhwa (2012) refer to as “knowledge networks.” But this is not the kind of knowledge most valued and rewarded by the academy. Academia values intellectual contributions that take the form of published scholarship. Traditional products of scholarship include academic journal articles, books and book chapters, and conference papers that employ systematic analysis and robust research methods. These traditional forms of scholarship are designed to advance and build upon prior knowledge, introduce or provide evidence for new theories or frameworks, and improve policy and practice. Improvement science does not readily yield traditional products of scholarship and remains challenged to find its place in academia. In terms of historically valued products in academia, the incentives are limited for scholars to partner with practitioners to engage in improvement science. Knowledge generation, at least in the academic sense, is privileged over practical attempts at performance improvement.

Improvement science thus suffers from a lack of legitimacy in an academic world that values knowledge generation through traditional research design. In order to gain legitimacy in academia, researchers engaged in improvement science will have to find ways to record and disseminate the *process* of improvement science. This would involve documenting responses to recursive inquiry cycles that occur in situ. In terms of higher education's responsibility, universities must be open to rewarding this type of engaged scholarship. By doing so, the research enterprise will move closer to achieving its ultimate mission of improving the human condition.

References

[1] Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, *41*(1), 16-25.

[2] Broekkamp, H., & van Hout-Wolters, B. (2007). The gap between educational research and practice: A literature review, symposium, and questionnaire. *Educational Research and Evaluation*, *13*(3), 203-220.

[3] Brown, J. R. (2011). *The laboratory of the mind: Thought experiments in the natural sciences*. Routledge.

[4] Bryk, A. S. (2009). Support a science of performance improvement. *Phi Delta Kappan, 90*(8), 597-600.

[5] Bryk, A. S. (2010). Organizing schools for improvement. *Phi Delta Kappan*, *91*(7), 23-30.

[6] Bryk, A. S. (2015). 2014 AERA Distinguished Lecture: Accelerating how we learn to improve. *Educational Researcher, 44*(9), 467–477.

[7] Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015a). Breaking the cycle of failed school reforms: Using Networked Improvement Communities to learn fast and implement well. *Harvard Education Letter, 31*(1), 1-3.

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

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

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

[11] Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, *32*(1), 5-8.

[12] Dynarski, M. (2015). Using research to improve education under the Every Student Succeeds Act. *Economic Studies*, *1*(8), 1-5.

[13] Ginsburg, A., & Smith, M. S. (2016). Do randomized controlled trials meet the "gold standard". *American Enterprise Institute. Retrieved March 18*, 2016.

[14] Junghans, T. (2018). "Don't Mind the Gap!" Reflections on improvement science as a paradigm. *Health Care Analysis, 26*(2), 124–139.

[15] Kolko, J. (2010). Abductive thinking and sensemaking: The drivers of design synthesis. *Design Issues, 26*(1), 15-28.

[16] Lewis, C. (2015). What is improvement science? Do we need it in education? *Educational Researcher*, *44*(1), 54-61.

[17] Oha, E., & Reeves, T. (2010). The implications of the differences between design research and instructional systems design for educational technology researchers and practitioners. *Educational Media International*, *4*(47), 263–275.

[18] McKenney, S., & Reeves, T. C. (2013). Systematic review of design-based research progress: Is a little knowledge a dangerous thing? *Educational Researcher*, *42*(2), 97–100.

[19] Nuthall, G. (2004). Relating classroom teaching to student learning: A critical analysis of why research has failed to bridge the theory-practice gap. *Harvard Educational Review*, *74*(3), 273-306.

[20] Phelps, C., Heidl, R., & Wadhwa, A. (2012). Knowledge, networks, and knowledge networks: A review and research agenda. *Journal of Management*, *38*(4), 1115-1166.

[21] Rohanna, K. (2017). Breaking the "adopt, attack, abandon" cycle: A case for improvement science in K–12 education. *New Directions for Evaluation*, *2017*(153), 65-77.

[22] Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change*, *7*(4), 221-258.

[23] Tiberghien, A., Vince, J., & Gaidioz, P. (2009). Design-based research: Case of a teaching sequence on mechanics. *International Journal of Science Education, 31*(17), 2275-2314.

[24] Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development, 53*(4), 5-23.

[25] Wenger, E. C., & Snyder, W. M. (2000). Communities of practice: The organizational frontier. *Harvard Business Review*, *78*(1), 139-146.