Carnegie Perspectives

Getting Ideas into Action: Building Networked Improvement Communities in Education

Anthony S. Bryk
Carnegie Foundation for the Advancement of Teaching

Louis M. Gomez
University of Pittsburgh & Carnegie Foundation for the Advancement of Teaching

Alicia Grunow
Carnegie Foundation for the Advancement of Teaching

GETTING IDEAS INTO ACTION: BUILDING NETWORKED IMPROVEMENT COMMUNITIES IN EDUCATION

This essay will be published in Spring 2011 in *Frontiers in Sociology of Education*, edited by Maureen Hallinan, Springer Publishing.

CARNEGIE PERSPECTIVES

Carnegie Perspectives is a daily blog that features handpicked news and commentaries on education. It also includes essays, updates, interviews, resources and connections to and about Carnegie's work. Carnegie Perspectives will also provide an in-depth look into our work in several areas, including Carnegie's approach to education research and development, our work in developmental mathematics, and information about a strand of work called Learning From Emerging Teaching Practice.

CREATIVE COMMONS LICENSE

ATTRIBUTION

REORGANIZING EDUCATIONAL RESEARCH AND DEVELOPMENT[1]

BACKGROUND

In a recent essay we argued that our current educational research and development (R&D) infrastructure fails to connect to enduring problems of improvement in our nation's schools and colleges (Bryk and Gomez 2008). An all too well-known sample of these problems includes: ethnically based gaps in academic achievement, too many adolescents dropping out of high school, too few children learning to read proficiently, and very low student success rates in our community colleges. We noted that educational problems like these continue to be vexing even though they have gained public policy attention and stimulated an extraordinary array of activity within the research and development community. Despite this activity, most assessments conclude that the R&D enterprise has not helped as much to date as one might hope and expect. A small but growing cadre of scholars and policy organizations have coalesced around an argument that the social organization of the research infrastructure is badly broken and a very different alternative is needed (e.g., Burkhardt and Schoenfeld 2003; Coburn and Stein 2010; Committee on a Strategic Education Research Partnership 2003; Hiebert, Gallimore, and Stigler 2002; Kelly 2006; National Academy of Education Report 1999).

In response, we argued for a more problem-centered approach that joins academic research, clinical practice and commercial expertise in sustained programs of Design-Educational Engineering and Development (DEED). We sketched out three overlapping phases of effective DEED, beginning with a set of alpha trials where a promising idea is attempted in a small number of places (Bryk and Gomez 2008; Bryk 2009). Extending this activity are beta investigations where DEED efforts deliberately focus on adapting the innovation so that it might be implemented with efficacy in more diverse settings. This, in turn, would lead to gamma-level activity that exploits evidence from large-scale use to continue to improve the innovation.

To engage such inquiries would require a profound shift in the social arrangements of R&D. Heretofore, nominal roles of researcher and practitioner have differentiated the arrangements of inquiry. Researchers, primarily those with PhDs in a cognate or applied discipline, did the

intellectual heavy lifting at the front end of the idea pipeline, while practitioners, those with on-the-ground experience, were expected to implement and adapt idealized innovations. Practitioners simultaneously engaged in local problem solving; however their efforts were rarely seen as significant in the infrastructure of educational R&D. The ideas seeded in our earlier essay and further developed here take a different perspective. We argue that the complex problems of practice improvement demand that a diverse mix of skills be brought to bear and require reconsideration of when and how in the arc of problem solving this diversity of expertise is best exploited. It demands new arrangements for disciplined inquiry where the work of research and practice join in a more dynamic and interactive fashion. It invites strong scholars to engage in applied R&D, but now in quite different ways in the pursuit of a science of improvement.

We detail in this essay how the social organization for such work might actually be carried out. Toward this end, we introduce the idea of a networked improvement community. We focus primarily on how research and practice communities might join in initiating such an enterprise. Our inspiration for the discussion below draws on insights from successful R&D activities occurring in diverse fields outside of education including the semiconductor industry, the Linux development community, and efforts at broad-scale quality improvements in health services. In each instance large networks have organized around complex problems and brought about remarkable change. Understanding these developments better, extracting core ideas, and translating them into more productive institutional arrangements for educational R&D pose important questions for learning scientists, organizational sociologists and political scientists interested in how expertise networks advance social improvement.

ORIENTING IDEAS

Improving the organization of educational R&D requires answers to three seemingly straightforward questions: First, what problem(s) are we trying to solve? Second, whose expertise is needed to solve these problems? And third, what are the social arrangements that will enable this work? While these questions appear to be simple, in the last decades our field's responses to them have been confused. When the answers to these questions are disorganized, the natural result is a cacophony of questions and innovations that fail to accumulate into real progress on core concerns.

Consider community college graduation rates, which is now a major concern for public policy. There is a broad public consensus that community colleges are an important opportunity resource for a large segment of America. In the research and development community, we recognize that improving community college graduation rates is an important priority. Is this level of agreement enough for sustained progress so that more students will successfully graduate? We argue no. Once problems like this cross some public policy threshold, a spate of uncoordinated research and development activity ensues. Some scholars might say community college failure rates are high because, "it is the textbooks"—we need to develop better and open source materials. Others

might say, "it is a student motivation problem"—let us organize learning communities to improve students' social connection to study groups across the college. Still others say "what we're teaching doesn't make sense when we look at students' educational and career goals"—so let us design/create more meaningful math courses. Still others say, "it is an institutional leadership failure"—let us get serious about leadership development initiatives. In short, for this and most other significant problems in education, there are many voices that attempt to characterize the problem.

We argue that large societal concerns such as improving community college success are complex problems composed of multiple strands (with numerous embedded micro-level problems) that play out over time and often interact with one another. More specifically, graduation rates in community colleges are an aggregate consequence of numerous processes such as courses taken, advising systems, course scheduling, etc. One does not improve graduation rates directly except by decomposing this big presenting problem into its constituent component processes, then analyzing the interconnections among them. It is within the problem system where students actually progress or fail.

Another question is, "Who should be doing the work?" If the listing of problem parts above captured even a small part of the problem ecology, then a very diverse colleagueship of expertise will be necessary to make progress (Bryk and Gomez 2008). Furthermore, these actors must be organized in ways that enhance the efficacy of individual efforts, align those efforts and increase the likelihood that a collection of such actions might accumulate towards efficacious solutions. While innovations abound in education, we argue that the field suffers from a lack of purposeful collective action. Instead, actors work with different theories of the same problem, activities are siloed, and local solutions remain local.

In this essay, we focus on an alternative social organization for this activity network: How might one structure and guide the varied and multiple associated efforts necessary to sustained collective action toward solving complex improvement problems? Drawing on Englebart (1992), we call this kind of organization a networked improvement community. We detail a set of structuring agents necessary for productive R&D to occur across such a community. We attend to how this form of social organization might come into existence and sustain participation over time in order to advance real improvements for significant numbers of students.

In an arena such as education, where market mechanisms are weak and where hierarchical command and control is not possible, networks provide a plausible alternative for productively organizing the diverse expertise needed to solve complex educational problems. Below we describe the organizing role that Networked Improvement Communities (NIC) might play here.

NETWORKS AS DESIGN COMMUNITIES

Networks enable individuals from many different contexts to participate according to their interests and expertise while sustaining collective attention on progress toward common goals. Organizational scholars have suggested that the novel interactions and information exchanges occurring within such networks make them particularly suitable for innovation and knowledge-intensive product design (Goldsmith and Eggers 2004; Podolny and Page 1998; Powell 1990). These more decentralized and horizontal work arrangements appear especially advantageous when, as von Hippel (2005) argues, "the problem-solving work of innovation requires access to 'sticky' information regarding user needs and the context of use." This knowledge is highly localized and thus costly to transfer. This latter consideration is especially significant in educational R&D where improving at scale requires coping productively with local diversity. The history of educational innovation is replete with stories that show how innovations work in the hands of a few, but lose effectiveness in the hands of the many (Gomez, Gomez, and Gifford 2010). At base here is a need for much better access to sticky knowledge. That is, we need design which explicitly aims to function in the hands of diverse individuals working in highly varied circumstances. We know all too well from past experiences that such contextual knowledge is not transferred easily across institutional lines to the academic labs or publishing companies where many educational tools and products currently are designed. In contrast, a network organizational approach can surface and test new insights and enable more fluid exchanges across contexts and traditional institutional boundaries—thus holding potential to enhance designing for scale.

NETWORKS AS LEARNING COMMUNITIES

The term network is used to describe a wide array of collectives. A networked improvement community is a distinct network form that arranges human and technical resources so that the community is capable of getting better at getting better (Englebart 2003). Englebart characterizes the work of organizations and organizational fields in terms of three broad domains of activity. In Englebart's terminology, A-level activity is the on-the-ground work of carrying out the organization's primary business. In the case of community colleges, A-level work is the front-line teaching and learning work of classrooms and includes units such as student support centers that offer tutoring services. Secondary or B-level activity describes within-organization efforts

that are designed to improve the on-the-ground work. In the community college realm, the work of institutional research units offers one example. These shops collect data about student success rates, and share that information with faculty and staff with the expectation that the data will inform subsequent improvements. C-level activity is inter-institutional, representing the capacity for learning to occur across organizations. Here institutions engage in concurrent development, working on problems and proposed solutions that have a strong family resemblance. Concurrent activity across contexts puts relevant aspects of the context in sharp relief and can help each local setting see its efforts from new vantage points. This is a boon to problem-solving. Englebart (2003) observes that C-level activity affords mechanisms for testing the validity of local knowledge, adjusting local understanding of the true nature of a problem, and advancing local support structures for improvement.

Applied inquiry in education has largely been about describing A-level activity, and on some occasions, evaluating it. Recently, we have seen a spate of interest in B-level activity, for example, in efforts to introduce evidence-based decision making as a guide to K-12 school reform. (See for example, Boudett, City, and Murnane 2005.) Likewise in community colleges, the Achieving the Dream initiative aims to develop local capacities to use data to inform improvement (see www.achievingthedream.org). This growing B-level activity is exciting because it lays the groundwork for the emergence of C-level inquiries and attendant possibilities for broad, inter-institutional social learning. Consequently, a more detailed account of how such networked improvement communities are initiated, organized and governed could be useful in efforts to enhance the overall productivity of educational R&D. Beginning work toward such an account is the aim of this paper.

EXEMPLARY NETWORKED IMPROVEMENT COMMUNITIES (NICS)

To help ground our conversation about how networked organizations can arrange themselves to accomplish improvement at scale in educational research and development, we have identified three, extant organizational cases that share a significant number of the features detailed in Englebart's analysis. We use insights from the International Technology Roadmap for Semiconductors (ITRS) in which (sometimes competing) organizations from across the semiconductor industry coordinated their innovation efforts. The shared road map was instrumental in catalyzing unparalleled R&D-based improvements in micro-electronics. We also look at Linux as a case of a loosely coupled collection of software professionals who volunteered their time to work cooperatively in an innovation network. Collectively they produced a complex and highly sophisticated modern multi-purpose operating system. And finally, we turn to the Institute for Healthcare improvement (IHI) that is creating a new ethos for how healthcare organizations work at global scale to continuously advance better healthcare outcomes. Each of these cases offers insights as to how a more effective educational R&D might be institutionally arranged. We have drawn eclectically on these experiences, and related theoretical accounts, to

detail framing elements of organizational structure, core work processes and operating norms for an educational networked improvement community.

A CASE OF LEARNING THROUGH DOING

As noted, reshaping educational R&D is a growing part of the contemporary scholarly and policy zeitgeist. This chapter contributes to this scholarship, and it has been informed by others. Only so much, however, can be learned through reflection; its natural complement is action to spur learning. That is, a concrete way to learn how a NIC might organize and carry out a better program of educational R&D is to build one. In this spirit, the Carnegie Foundation for the Advancement of Teaching has adopted a learning through doing orientation. Under the Foundation's umbrella, and in partnership with several other colleagues and institutions,[2] we are now initiating a prototype NIC aimed at addressing the extraordinary failure rates in developmental mathematics in community colleges. As noted earlier, multiple processes combine to create observed community college outcomes. We sought to initiate a NIC around a high leverage wedge into this organizational system. Practitioners and researchers now agree that a key contributor, arguably the most important contributor, to low graduation rates in community colleges is the high failure rates of students in developmental mathematics courses (see for example, Cullinane and Treisman 2010). Redressing this is a well-specified problem around which a NIC can organize. In the pages that follow, we describe the rationale and design for this NIC and draw on emerging practices within it to illustrate how our framing ideas about NIC are becoming manifest in action.

The nation's 1,000 community colleges enroll more than 6 million students or upwards of 40 percent of all postsecondary students in the United States.[3] These institutions are the front-line in our nation's efforts to advance social equity and supply labor needs for a 21st century economy.[4] At present however, many students enter community colleges with high aims and ambition, only to languish, sometimes for years, in developmental courses that are non-credit bearing and do not

---

[2] Of special note in this regard is the Foundation's partnership with the Dana Center at the University of Texas. Dana has lead design responsibility for developing the initial instructional kernel for Statway. This includes pathway outcomes, a modular structure for the curriculum, classroom lessons and assessments. In addition, the executive director of the Dana Center, Uri Treisman, also serves as a Carnegie senior partner. In this latter role, Treisman co-leads the policy outreach and planning for scale team in the network.

[3] American Association of Community Colleges; Retrieved from http://www2.aacc.nche.edu/research/index.htm on Sept 11, 2009.

[4] The magnitude of community colleges' collective responsibility nearly doubled in July 2009 when President Obama called for an additional 5 million community college degrees and certificates by 2020. To achieve this scale under a constrained timeframe requires bold innovation. Entitled the American Graduation Initiative, the plan as proposed will invest 12 billion dollars to invigorate community colleges across the United States by funding improvements in physical infrastructure, developing challenge grant mechanisms and creating a virtual course clearinghouse. Specifically the President highlights open, online education as a strategy for reaching more non-traditional students, accelerating students' progress, helping students persist, and improving instructional quality.

move them toward a degree, certificate or transfer. This is true especially in mathematics. Recent studies report that between 60 and 70 percent of students who are referred to developmental mathematics do not successfully complete the sequence of required courses.[5] Many spend long periods of time repeating courses or simply leave college. Either way, they do not reach their career goals.

A careful analysis of this larger problem reveals a complex of sub-problems operating within community colleges that contribute to the high failure rates. Instructional systems do not engage student interest in learning; student support systems inconsistently meet students' needs; human resource practices and governance structures create barriers for change; and there is insufficient access to data and insufficient use of data to inform improvements. Small gains may be possible by focusing on single elements, but dramatic change ultimately requires a systems view of how these elements (and others) inter-lock to create the overall outcomes currently observed.

The Carnegie Foundation has set out to catalyze and support the growth of a networked improvement community aimed at doubling the proportion of community college students who, within one year of community college enrollment, are prepared mathematically to succeed in further academic or occupational pursuits. Carnegie's first effort in this regard is to launch a Carnegie Statway Network.[6] This network seeks to redesign traditional developmental mathematics by creating a one-year pathway to and through statistics that integrates necessary mathematics learning along the way.

The first participants in the Statway network are nineteen community college teams comprised of three faculty members, an institutional researcher and an academic dean or vice-president. These teams are now working together with Carnegie to co-develop a set of base resources for the network. Faculty members will develop, test and refine an initial set of instructional resources. Common assessments and a lesson study methodology anchor their activity and set the stage for the continuous improvement of the instructional materials over time. The institutional researchers are working together to build common evidence systems to enable the network to measure, compare and improve the performance of Statway students both within and across institutions. The deans and administrators from each college are addressing the multitude of logistical issues that arise in embedding an innovative design within their institutional contexts.

The work of these teams is supported, in turn, by expert others. As these pilot efforts proceed, the network will address concerns around faculty development and where and how technology

---

[5] "Referral, Enrollment and Completion in Developmental Education Sequences in Community Colleges," Bailey, Jeong, and Choo, CCRC Working Paper, no. 15 Dec 2008, (revised April 2009). These data were obtained from Achieving the Dream campuses and compared to NELS 88 data.

[6] It is argued more generally that we need a small number of more structured pathways to success. Statway is Carnegie's first effort in this regard. The Foundation also will support efforts on a second pathway, called Quantway, seeking to achieve similar goals for students with somewhat different career aspirations.

can add value. The network will form a robust information infrastructure to inform continuous improvement. It must consider how issues of literacy and language mediate mathematics learning, and scrutinize how the vast array of extant academic, social, psychological/counseling services can be better integrated to advance student success. These are all key to advancing efficacious outcomes reliably at scale. Taken together, this assembled expertise provides the initiating social form for our NIC, which we call a Collaboratory. As the network evolves, Collaboratory membership will expand to other specialized practitioners, design-developers and researchers as new needs and priorities come into focus.

The Statway design products co-developed within the Collaboratory will belong to a growing network improvement community and serve as base resources for the network to further improve over time. Involvement in the network also will advance participants' instructional and institutional expertise, thereby creating a cadre of leaders and champions for subsequent expansion of the network. Any intervention that is human and social resource intensive, as is the case for most educational improvement efforts, requires organizational and institutional structuring to build capacity. As such, attending to how to engender a proper organizing structure for problem solving in the alpha stage is a key issue for activity expansion into the beta phase (Bryk and Gomez 2008). We now turn to the issue of network structuring.

STRUCTURING AGENTS

All networks have rules and norms for membership. They maintain narratives that detail what they are about and why it is important to affiliate. In one way or another, networked communities make clear who is allowed to join, how to join, and how to participate. Membership criteria may be very loose and broad. In a community like Facebook, for example, literally everyone can join. (There are, of course, within Facebook, many sub-communities with restrictions.) Facebook is essentially an open community.

Open networks abound in education. In the main, they function as free-floating idea bazaars, contexts for self-expression, and places to share information. In fact, the current social organization of educational R&D functions much like an open network. It is characterized by a multitude of voices lobbying for preferred approaches, but with weak mechanisms for directing intentional action that cumulates in coherent solutions to complex problems. In this regard, educational R&D's inability to orchestrate such improvement is akin to a market failure.

In contrast, a networked improvement community is an intentionally formed social organization. Its improvement goals impose specific demands on the rules and norms of participation. We detail here a set of structuring agents necessary to form participation in such an intentionally designed network so that coordinated R&D can occur on a focal problem. "Getting these agents

right" is key to unleashing individual creativity, while also advancing joint accountability toward collective problem solving.

COMMON TARGETS AND MEASUREABLE AMBITIOUS GOALS

The community of practice has become a prevalent organizational arrangement in education to support collaboration.[7] Communities of practice require that members have interests in common. For example, a community of practice devoted to teaching high school biology through open-ended and long-term project investigation, centers its activities on sharing ideas about ways to accomplish projects more effectively in classrooms.[8] While communities of practice may form around a common concern, such as improving the execution of science projects in biology classes, their goal is to support individual action. In the best of cases, communities of practice may share some common artifacts, such as a rubric that specifies elements of science projects (e.g., developing a driving question). Rarely do these specifications, however, lead to the execution of common work, to shared outcome measures and to mechanisms for comparing results by which progress toward specified goals can be judged. Coordination in a community of practice is limited to maintaining a social focus on a common problem, akin to keeping members in the same idea-ballpark.

In contrast, we posit that a networked improvement community requires more structured social arrangements. Participants in a NIC endorse shared, precise, measureable targets. Participants agree to use what is learned, from working toward meeting the targets, to setting new targets aimed at ever more ambitious goals. In this regard, shared measureable targets help a community stay focused on what matters, from the community's perspective. They catalyze discussions among participants as to why we should attend to this rather than that. They demand argument about what is likely to afford more immediate progress. In this regard, they introduce some discipline in priority setting as it interacts with an individualistic rhetoric of "I am interested in…" NICs rebalance arguments from personal interests to targets. They also shift the location of goals from the personal "I'" to the collective "we".

The semiconductor industry provides an illustrative example. Gordon Moore, co-founder of Intel, noted in 1965 that the number of transistors that could be placed cost-effectively on an integrated circuit had doubled every year since the invention of the transistor. Moore saw no technical reason why this trend would not continue for at least the next twenty years. Moore's prediction turned out to be correct, and his observation was later named Moore's Law.

In the semiconductor industry, Moore's Law is a beacon. It guides work for a diverse collection of colleagues within, and across, firms in that industry. It shapes the activities of engineers who design and construct devices and it shapes how corporate leaders invest capital. Further, since

---

[7] For a seminal text on this topic see Wenger (1999).
[8] See for example Ruopp et al. 1993; Schlager, Fusco, and Schank 2002.

Moore's Law is anchored in evidence about past performance and a perspective of feasible developments, it offers reason to believe that stretch targets are actually attainable. The combination of feasibility, and the knowledge that everyone is working in a common direction, can have significant disciplining power in a community. In essence the targets help create virtuous cycles of joint accountability.

This feature of targets also is visible in organizing the efforts of the Institute for Healthcare Improvement (IHI). In each IHI initiative, explicit attention focuses on specifying precise, measureable goals for each improvement. Participants work under a shared understanding that, "some is not a number and soon is not a time."[9] Defining measureable outcomes and timelines to achieve those outcomes guides efforts in IHI's improvement communities.

Targets have at least one additional important benefit. They engender ongoing vetting processes. Targets are under constant negotiation in networked improvement communities. Take the case of Wikipedia. Most Wikipedia users think of it as a reference product. For its members, however, it functions as an argument platform. The peer-to-peer platform is a vehicle that structures and propels their conversation. The online encyclopedia is simply the very useful emergent product of all that talk (Shirky 2008). In a similar way, the act of setting common targets in network improvement communities is a way for community members to vet goals and sharpen shared understandings. The process draws people into regular conversations that develop into distinct communication forms that then structure behavior. Consequently, evolving targets are more than just a way to get to a product. The evolution is a process that, in and of itself, shapes and strengthens activity in a community.

The importance of targets has not been lost on educators. For example, it is a core element in the No Child Left Behind Act (NCLB). NCLB established a measureable goal of 100 percent proficiency on state tests in math and reading by 2014. This explicit target has, in many ways, encouraged the sort of behavior we might expect, given our previous discussion. In light of NCLB's target, learning standards were revised by professional societies, states and cities across the country to align with the target. The assessment industry went into high gear to build tests to help states judge whether schools and students were meeting, or on track to meet, the standards. At the same time researchers and designers pondered and piloted new assessments that might be better at judging progress. States and districts invested in data warehouses to report performance data that highlights, in granular detail, who is and who is not making average yearly progress (AYP) toward targets. School leaders focused attention and resources on disadvantaged children, especially those just below proficiency levels, as their progress was key to achieving AYP

---

[9] See for example the overview of IHI's 5 Million Lives campaign:
http://www.ihi.org/IHI/Programs/Campaign/Campaign.htm?TabId=1

benchmarks. An industry of supplementary support services, especially individual student tutoring, grew rapidly as well. In short, a flurry of activity accrued in the wake of NCLB targets.

What did not cohere around NCLB was a full-fledged, networked improvement community. In comparison, the semiconductor industry took great care in creating targets that were viewed as attainable; whereas NCLB 2014 reading and math targets generated great skepticism.[10] For NCLB, there was no disciplining equivalent to Moore's Law. That law was anchored in empirical evidence about what had been achieved previously, combined with a shared field perspective that further improvements along these lines actually might be attainable. In contrast, NCLB goals represented an expression of valued social aims imposed by legislative action. No empirical evidence existed from past practice that the goals could be achieved and no community formed around their continued elaboration and refinement. To be sure, NCLB motivated individual actions, and many goods and services were purchased in an effort to reach the targets, but accumulating R&D for improvement was never vitalized.

These experiences have important implications as we think about targets in the context of Statway. We too lack the luxury of a disciplining framework like Moore's Law, and worry that imposing socially valued outcomes by fiat lacks the organizing power needed. While we recognize the power of targets, we also understand that they must be valued and considered attainable by a community.

To begin a redress to this concern, each Collaboratory college will establish a performance baseline for students eligible to be served by Statway. This baseline will include common measures of student learning and program progress. In addition we will collect data that refine our understandings about the student population being served; for example, their math course-taking history and proficiency, their language and literacy background and proficiencies, their motivation and determination to succeed in community college, etc.[11] Performance college-by-college in subsequent years will be judged against their local baselines. A distribution of effects will naturally emerge as the network accrues results from multiple sites implementing and refining Statway over several years. These results likely will vary from null findings in some situations to quite substantial improvements in others. We anticipate that somewhere along this

---

[10] For example, Robert L. Linn of the National Center for Research on Evaluation is widely quoted as saying of NCLB: "There is a zero percent chance that we will ever reach a 100 percent target." (http://www.washingtonpost.com/wp-dyn/content/article/2007/03/13/AR2007031301781.html). Also see Bryant et al. 2008.

[11] This population definition process is now underway. It includes measures from student math and reading placement tests, and English language capabilities.

distribution, say for example at the 75th percentile, a sense of feasible stretch goals should come into view.[12]

At base, two elements are key to establishing feasible targets that are generative toward improvement at scale. First, the variability in results achieved in the Statway network will be public to the Statway participants. And second, there will be a collective agreement to use results to continually refine targets in order to insure community ownership. As occurred in the semi-conductor industry, we posit that a joint accountability dynamic will emerge through this process of reviewing network-wide results. Knowing that others engaged in the same endeavor may be achieving at higher rates creates incentives for learning how these successes are occurring. That is, as the network focuses on a comparative analysis of results for its ongoing target setting, the same processes also function to incent individual learning and improvements network-wide.[13]

MAPPING A COMPLEX PROBLEM-SOLUTION SPACE: FORMING A SHARED LANGUAGE COMMUNITY

In addition to shared targets, mechanisms also are needed for coordinating efforts across diverse individuals and organizations engaged in a marketplace of parallel activity. The semiconductor industry uses an artifact called a roadmap to specify how targets become realized in the work of design, development and engineering in different contexts. The terrain for possible innovation is vast and complex. The roadmap organizes the challenges to be confronted in this space in agreed upon ways. It establishes standards for how developments in different domains must fit together, and then sets micro-targets, domain by domain. In these ways the roadmap helps to coordinate the activity so that different innovations in hardware and software can be expected to inter-operate at designated times in the future. In the industry's view, road mapping is a "practical approach to deal with the complex process of technological innovation" (Schaller 2004:13).

In essence, the roadmap reshapes accountability relationships that go beyond the confines of an individual firm or laboratory (Schaller 2004). Participants, both individuals and firms, are not autonomous actors operating within a disconnected marketplace. Rather, they form a densely connected network of peers who share a focused interest in common regions of the roadmap. Of note, today's technological climate includes the pervasive use of peer-to-peer collaboration tools,

---

[12] Ideally, the Collaboratory would be able draw results from previous institutional improvement efforts in the general domain of developmental mathematics education to set network wide goal. Absent the shared empirical discipline, such common data structures do not currently exist in the field. In contrast, were a community to embrace PDSA cycles as a common inquiry, see following section, such data might exist in the future. For an example of such a database in K-12, see research by Consortium on Chicago School Research. Bryk et al. (2010), for example, document rates of learning improvement across more than 400 elementary schools during a six-year period. These results provide an empirical basis setting improvement standards. Specifically, we know that improvements in annual learning gains of 10 percent or more in reading and 20 percent or more in mathematics are attainable.
[13] For an example in healthcare improvement, see Gawande's (2007) account of improvements in the treatment of cystic fibrosis across a health center network.

where individual activity is rapidly shared and transparent. For example, if network members working on common problems hear that others in the community have reached specified performance targets, it may spur the community to speed up work on competing products or push forward more rapidly on new products given these reports.

For these reasons, the process of mapping the space for innovation development strikes us as another critical structuring agent for an educational improvement network. Problems such as dramatically improving student success in developmental math are not simple. Multiple processes happen simultaneously, and multiple sub-systems within a community college are engaged around them. Each process has its own cause and effect logic, and these processes interact with one another over time to produce the overall outcomes we observe. Put simply, the extraordinary high failure rates in developmental mathematics in community colleges is a complex problem system.

The intrinsic complexity of such problem systems means that most participants appreciate only the parts of the system that seem particularly relevant to their role. Absent a working theory of the whole, interventions fail because of externalities not considered (also known as implementation failures), even though these are often predictable. For example, curriculum interventions often fail because of inadequate professional development. At base here is a natural human tendency to grasp for promising solutions or best practices without fully understanding how such solutions must be integrated with other solutions and pre-existing organizational conditions.

In short, for a NIC to make headway towards constructive improvements on a complex problem, the community needs to detail the contours of its problem-solution space. Similar to the semiconductor industry, this includes elaborating various elements or sub-systems that form it and the inter-relations among them. Key here is to "carve the system at the joints" so that independent work can occur on pieces of the system and so that these components can be aggregated into more systemic solutions.

An important aspect of tools like the roadmap is that they help people see the challenges of innovation jointly and from multiple perspectives. For example, the semiconductor roadmap carved the problem at the level of individual devices while at the same time specifying the inter-operability of the parts. This problem decomposition was then coupled with a temporal dynamic. That is, the roadmap also showed targets for technical performance, as they should unfold over time. The end product of such road mapping is a common language for organizing the diverse efforts occurring within a design and development community. Highly independent activities may occur across time and space, but the overall endeavor now coheres. Of note, the roadmap also provides the natural framework for accumulating field knowledge as it is developing. Over time, the roadmap is a persistent indicator that these joint efforts amount to tangible and

important progress. In short, mapping the problem-solution space is key for coordinated work to occur and for improvement knowledge to accumulate.

PROGRAM IMPROVEMENT MAPS

We have developed two tools to assist efforts to decompose the problem-solution space confronted by the Statway Collaboratory. The first we call a program improvement map. The map seeks to align a network around a common understanding of the problem at hand. While decomposing a complex improvement problem into component parts, it also highlights the character of the system that embeds it. The anatomy of a problem is further detailed in terms of interacting subsystems, specified targets by domain, and the particular audiences for whom these outcomes are especially relevant. The map specifies the elements in how people currently work together to produce observed outcomes, and in so doing, organizes the challenges that must be confronted if substantial improvement is to occur.

The challenges faced by community colleges in seeking to use statistics as the curricular vehicle for revamping developmental math is by no means just an instructional system problem. While knotty curricular and pedagogical problems must be unraveled, to be successful, Statway also must reach deeply into other institutional aspects of the community college and the policy infrastructure that surrounds it. The program improvement map encourages us to think through a detailed characterization of how these system elements operate in tandem with one another to produce the overall outcomes currently observed. The map also puts into relief how the efforts of other organizations beyond the colleges themselves, like curriculum providers and assessment developers, contribute to these outcomes. In so doing, it brings into focus how their efforts join the challenge space for innovation. In short, the program improvement map, like the roadmap, is a coordination device for diverse actors. It seeks to keep the improvement priorities of a network and their interconnections in explicit view as participants work on different parts of the problem.

Figure 1 offers a prototype program improvement map for guiding network activities in the Statway Collaboratory.[14] At first glance it is apparent that this map aims to convey a systems perspective. The "challenge space" seeks to identify the organizational elements that need to be addressed in striving toward the targeted outcomes for students. For example, the map recognizes that improvement poses challenges for both the instructional and the human resource systems. This recognition is meant to stem the competing interventions problem that often results

---

[14] Our intent here is not to argue for the adequacy of the specific details offered, but simply to illustrate the system character of a problem and how it might be "carved at the joints" to guide subsequent efforts.

when policymakers, researchers, and practitioners gravitate toward one or another reform idea, believing that this is the silver-bullet solution. The program improvement map highlights the fact that there aren't any simple solutions. It documents how any specific solution likely will touch many other aspects of the problem space well beyond the confines of its own box. Consequently, the adaptive integration of a component solution with its larger organizational context is now placed firmly on the design table.

Figure 1. An Illustrative Program Improvement Map
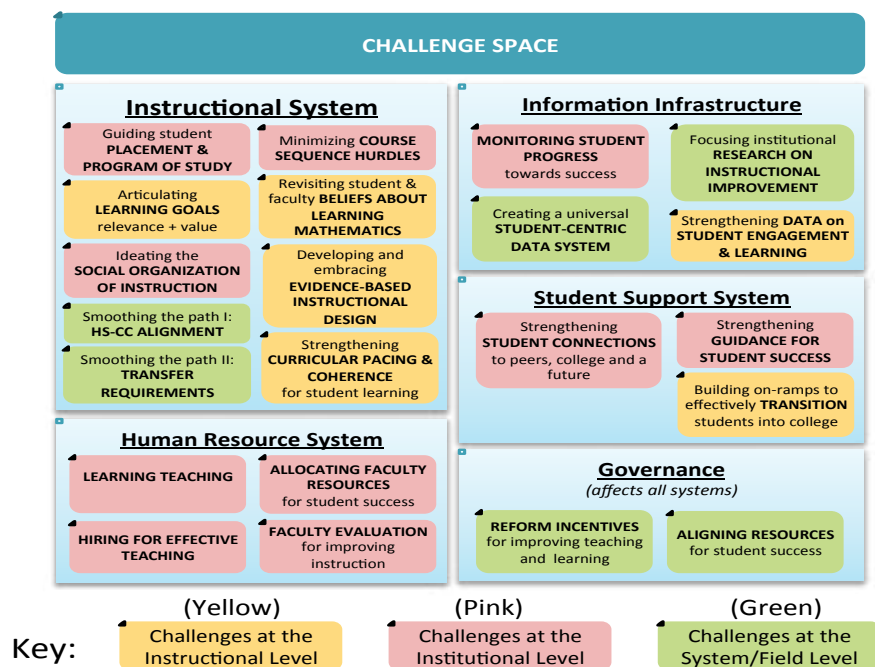
In short, the purpose of a tool like a program improvement map is to provide an end-to-end description of the challenge space. It encourages members of a networked improvement community to locate specific interventions in the larger problem space and begin to anticipate and problem solve around the systemic inter-connections of any intervention. In this light, an intervention is, in essence, a hypothesized solution path through the program improvement map's space.

DRIVER DIAGRAMS

Complementing the program improvement map is a second tool—the driver diagram. Drawing on a practice from improvement science (Langley et al. 1996), a driver diagram encourages network actors to explicate causal thinking; that is, how a proposed solution path responds to
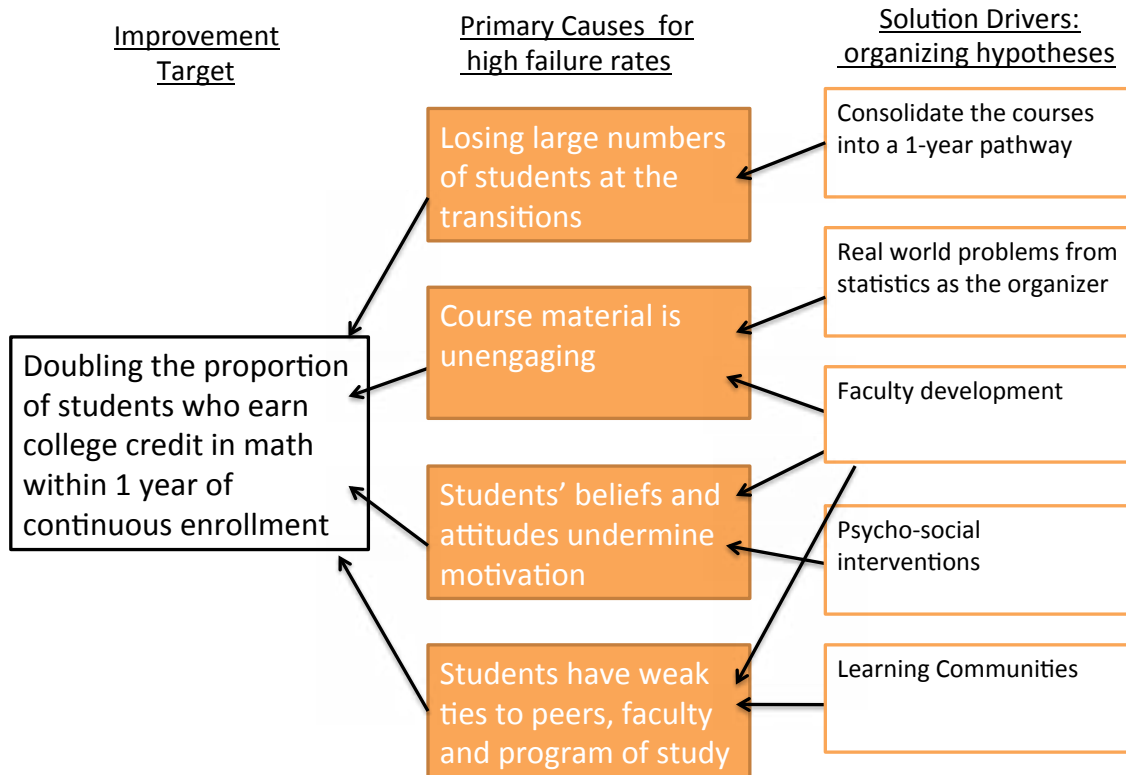
current understandings of the problem. The driver diagram requires attention to the specific hypotheses undergirding improvement solutions. These hypotheses are open to explicit study with common modes of inquiry.

In general, a driver diagram has three key elements: targets, primary drivers, and secondary drivers. The target is one of the community's agreed upon outcomes from the program improvement map. The primary drivers are the major causal explanations hypothesized to produce currently observed results. Secondary drivers, in contrast, are interventions in the system aimed at advancing improvement toward targets. Any argument for a specific secondary driver must explicate thinking about how a proposed intervention interconnects with understandings about primary causes or primary drivers for the outcomes currently observed. In so doing, an explicit causal explanation of problem-solution is developed. This can then be tested and refined against evidence.

Figure 2 illustrates a simple driver diagram. We begin with a specific target: Doubling the proportion of students, p, who begin in developmental mathematics and achieve college credit in this subject within one year of continuous enrollment.[15] In this illustration, four primary drivers are hypothesized. The first focuses on problems associated with student course transitions. We know that we lose large numbers of students at transition points; for instance, when they complete one developmental course but may not enroll in the next. (Indeed, for many students, three to four developmental math courses may be required before reaching a credit-bearing class, and this may take two or more years to complete.) Seeing the problem in this way led us to a specific change proposal: consolidate the pathway to one year of intensive instruction and have students enroll in a single pathway rather than separate courses where they have multiple opportunities to fall off the tracks. The latter is called a secondary driver.

---

[15] Given the space limits of the paper, we have constrained the example to a very rudimentary exposition. Our intent is simply to illustrate the tool rather than argue the merits of this particular instantiation.

Figure 2. Illustrative Driver Diagram for One Statway Aim

Improvement Target	Primary Causes for high failure rates	Solution Drivers: organizing hypotheses
Doubling the proportion of students who earn college credit in math within 1 year of continuous enrollment	Losing large numbers of students at the transitions	Consolidate the courses into a 1-year pathway
	Course material is unengaging	Real world problems from statistics as the organizer
		Faculty development
	Students' beliefs and attitudes undermine motivation	Psycho-social interventions
	Students have weak ties to peers, faculty and program of study	Learning Communities

Typically a driver diagram includes multiple inter-related hypotheses about the presenting problem and plausible solutions.[16] Figure 2 offers three additional examples of primary drivers: the problem of unengaging course materials; students' beliefs that they are not good at math; and weak social ties that do not strongly connect students to peers, to faculty, and to a specific program of study. In turn, each of these primary drivers links to a specific secondary driver or hypothesized improvement intervention. To solve the problem of unengaging course materials, for example, we might use real world concerns and analysis of relevant data as the backbone of instruction. To help students develop a stronger self-concept, we might introduce short-term psycho-social interventions that have proven to change student beliefs about efficacy of personal effort. And finally to develop more robust social ties, we might consider whether interventions such as learning communities might provide a vehicle for creating social capital for success. More generally, primary drivers may link to multiple interventions or secondary drivers.

---

[16] We note that in use the adequacy of a driver diagram is subject to empirical test. If all of the primary drivers have been identified, and an organization demonstrates change on each, then measureable improvements on the specified targets should occur. If the latter fails to materialize, some aspect of the driver diagram is underspecified. At base here is an organizing idea in science. Measurement and theory development move hand-in-hand. Theory sets out what we should measure; measurement in turn forces clarification on theory.

Likewise, a secondary driver also may respond to multiple primary drivers. For example, explicit attention may focus a faculty development initiative, a secondary driver, on both the importance of relational practices that sustain student engagement in instruction and the limited expectations that faculty may convey inadvertently about students' possibilities for success.

In sum, explicit problem decomposition coupled to explicit causal logic in intervention design is a critical agent guiding activity across a networked improvement community. Program improvement maps and driver diagrams are two promising, analytic tools that could be used to support coordinated work. Intellectually powerful forces anchored in personal belief and role-specific experiences tend to direct garden-path interveners away from systems thinking and toward silver-bullet solutions. A tool kit that includes program improvement maps and driver diagrams can discipline a community of interveners to see problems with larger, common eyes even as they may intervene in very specific ways. Regardless of whether the particular tools introduced here are actually used, we posit that the basic functions served by them must be addressed in some fashion for an improvement network to learn.

COMMON PROTOCOLS FOR INQUIRY

Effective network action also requires common protocols that allow participants to share, test, and generalize local learning across a professional community of practice. As Hiebert et al. (2002) have noted, such common inquiry protocols distinguish activity aimed at building professional knowledge from individual clinical decision- making.

We posit that a networked improvement community must engage in a disciplined approach to inquiry. For this purpose, we draw from core principles in continuous improvement pioneered by Deming, Juran and others (e.g., Deming 2000; Juran 1962). We especially are indebted to insights gleaned from extensive use of these principles in advancing healthcare services improvement. Like education, health services are carried out through complex organizations. (Think of a hospital or a network of hospitals being comparable in complexity to a community college system that spans multiple campuses or a school district.) Like physicians, school and college faculty expect to have discretion to determine how best to respond to a particular set of presenting circumstances. Both enterprises are human and social resource intensive, and both operate under largely decentralized governance arrangements while also being subjected to increasing external regulation. [17]

Research on health services improvement has surfaced recently in popular accounts such as Gawande's *Better* (2007) and more recently the *Checklist Manifesto* (2009). A dynamic group of

---

[17] Of note, the primary mode of inquiry in this applied domain does not typically follow the clinical trials methodology that characterizes the development and marketing of new drug treatments. The more common protocol involves establishing a baseline of results and comparing subsequent performance against this baseline. See, for example, Gawande (2007; 2009).

leaders has been building this field for over two decades.[18] Particularly noteworthy are the efforts of Berwick and colleagues (2008) at the Institute for Healthcare Improvement (IHI) who pioneered a set of inquiry practices and conceptual frameworks that are now broadly applied to improving health services worldwide.

A core set of principles undergirds this work and forms a science of improvement[19] (Berwick 2008). As is customary in scientific inquiry, common protocols discipline the work carried out by individual participants. These protocols guide local efforts to introduce changes and examine whether these changes actually are improvements. This is akin to Englebart's "B-level learning activity" mentioned earlier. Simultaneously, these protocols also structure possibilities for accumulating evidence from diverse inquiries occurring across varied contexts and time. They afford data for examining the replicability of results, as is the focus of a meta-analysis. Even more important, the breadth of evidence generated, coupled with diversity among network contexts and participants, creates opportunities for new synthetic insights to arise that are unlikely to occur within any one study. In short, these common protocols operate as structuring agents for the systematic inter-organizational learning that characterizes the C-level activity detailed by Englebart. Such learnings are largely missing in educational R&D at present. Instead, we live stuck between two polar views. On the one hand, a robust infrastructure has emerged for examining narrow, focused propositions through large, randomized field trials. On the other hand, there is a long tradition in education of local learning from the actions of individual practitioners. In the following we discuss these two traditions of translational and action research and argue for a third way.

In its idealized form, translational research envisions a university-based actor drawing on some set of disciplinary theory, such as, learning theory, to design an intervention. This activity is sometimes described as pushing research into practice (see, for example, Coburn and Stein 2010:10). After an initial pilot, the intervention is then typically field tested in a small number of sites in an efficacy trial. If this proves promising, the intervention is then subject to a rigorous randomized control trial to estimate an overall effect size. Along the way, the intervention becomes more specified and detailed. Practitioner advice may be sought during this process, but the ultimate goal is a standard product to be implemented by practitioners as designed. It is assumed that positive effects will accrue generally, regardless of local context, provided the intervention is implemented with fidelity.

In contrast, action research places the individual practitioner, or some small group of practitioners, at the center. The specification of the research problem is highly contextualized and the aim is localized learning for improvement. While both theory and evidence play a role,

---

[18] For a very readable historical narrative on this account, see Kenney (2008).

[19] Throughout this essay we use interchangeably the terms science of improvement and improvement research. One or the other may be more connotative depending upon context and audience.

the structures guiding inquiry are less formalized. Common constructs, measures, inquiry protocols and methods for accumulating evidence typically receive even less emphasis. The strength of such inquiry is the salience of its results to those directly engaged. How this practitioner knowledge might be further tested, refined and generalized into a professional knowledge, however, remains largely unaddressed (Hiebert et al. 2002).

A science of improvement offers a productive synthesis across this research-practice divide. It aims to meld the conceptual strength and methodological norms associated with translational research to the contextual specificity, deep clinical insight and practical orientation characteristic of action research. To the point, the ideas sketched below are consistent with the b basic principles of scientific inquiry as set out by the National Research Council (Shavelson and Towne 2002). [20]

A CONTINUOUS IMPROVEMENT ETHIC ENGAGED ACROSS A NETWORK

Shared narratives integrate collective experience. The main theme in the narrative for an improvement network is *Learning through Doing*.[21] Multiple cycles of design-engineering-development characterize the improvement efforts occurring within a participating classroom, college, or individual commercial firm.[22] In principle, each cycle propels some bit of local learning. When parallel development activities occur in different sites at the same time, a network can learn from the ensemble of these experiences. This increases the overall odds of efficacious outcomes emerging more reliably at scale. This practice of learning through doing enlivens the mantra of continuous improvement that deficits are a treasure.[23] Each process failure provides an opportunity to learn and to improve both locally and network wide.

Cronbach sketched out this approach to the social organization of applied research over 30 years ago. Cronbach (1980) argued that sturdy evidence to inform improvement at scale is more likely to arise out of a fleet of studies rather than one big field trial. Although the causal warrant for

---

[20] Shavelson and Towne (2002) identify six core principles. These include: specific questions to be investigated empirically; theory guides the investigation and generating cumulative knowledge is a goal; use of methods that permit a direct investigation of the question; a coherent and explicit chain of reasoning; efforts to replicate findings across a range of time and places and synthesize and integrate results; and open research to scrutiny and critique where objectivity derives from the enforced norms of a professional community. All of these are operationalized across an improvement research network.

[21] For a classic exposition of these ideas see Lewin (1942).

[22] The ideas developed in this section apply equally to all participants in a network improvement community. Depending upon the particular improvement objective, the units of interest might be individual classrooms, study centers within community colleges, departments or entire colleges. They also apply to commercial firms developing new tools, goods and services for this marketplace. In the interest of simplicity, we use the term colleges as a placeholder for this larger and more varied domain of participants.

[23] http://www.ihi.org/IHI/Topics/Improvement/ImprovementMethods/ImprovementStories/ TreatEveryDefectasaTreasure.htm

results in any one small study may well be weaker than the standards espoused for a rigorous clinical trial, a fleet of coordinated inquiries can generate much richer information about how an innovation actually functions when diverse participants are working in varied organizational contexts and time periods. The latter is essential knowledge for achieving efficacious outcomes more reliably at scale.[24]

This idea reminds us of the two spans of the inference problem identified by Cornfield and Tukey (1956). Getting a precise estimate about a treatment effect in a fixed setting, as for example in some non-randomly selected set of sites studied in an educational randomized control trial, takes us only part way from data to practical inference. One must still negotiate a second span, which is the capacity of such data to actually guide improvement.[25] As Cornfield and Tukey point out, the two spans sit in some tension with one another and useable research entails effective compromise.[26] Applied research in education today has become hyper-concerned with the internal validity of the individual field trial. Treatises on modern causal inference place primacy on the word cause while largely ignoring concerns about the applicability of findings to varied people, places and circumstances.[27] In contrast, improvement research must take this on as a central concern if its goal is useable knowledge to inform broad scale change. This consideration has important implications, as we will elaborate further.

THE CENTRAL ROLE OF PRACTICAL MEASUREMENT

While individual practitioners may rely on personal observations for their learning, improvement at scale entails common measurement.[28] The latter is key to learning across a network from the natural variation arising within it. Such measurement includes longer-term outcomes, both intended and unintended. It also requires attention to process measures and shorter-term effects

---

[24] A close parallel to this in healthcare is the idea of complex treatment regimes and how the multivariate evidence needed to inform this from the more univariate evidence of average treatment effects generated by RCTs. For a good example see the *Patients Like Me* web site (http://www.patientslikeme.com/) as a knowledge base for chronic care. Patients have individual treatment histories and may be involved in multiple therapies simultaneously. Data to inform "what is right for me" involves more complex information structures than the on-average results derived from randomized control trials of individual therapies.

[25] Formally, Cornfield and Tukey (1956) use the term inference meaning how one might apply the results of an experiment to a larger and different set of cases. Modern causal inference places primacy on the word "cause" and not the idea of "generalization." The latter in contrast is key to Cornfield and Tukey's argument.

[26] See also the classic distinction between internal and external validity introduced in Campbell and Stanley (1963) and further elaborated in Cook and Campbell (1979).

[27] See Weisberg (2010) for an explication of this argument.

[28] Note, we focus here on the common core of data that regularly informs the work of NIC participants and provides one basis for cross-network learning. A sub-network within a NIC can, of course, also engage in specialized individual studies and one-time field trials. In fact, we are organizing as part of Statway an "alpha lab" that would bring an expanding array of applied researchers into this problem solving research. The initial agenda for the alpha lab will focus on opportunities to deepen students' mathematics understandings, strengthen motivation for sustained work in the Pathway, and address literacy and language demands in statistics instruction.

on students. For example, in our developmental mathematics education network, we will track long-term outcomes such as the percentage of students who successfully complete a college level math course, or eventually earn an AA degree or transfer to a four-year institution. These are key summative measures, but they also tend to operate as lagging indicators. If some important process changes are affected, a jump in these aggregate indicators may accrue a year or two later. In general, real process improvements manifest in lagging indicators sometimes well after the actual improvements have occurred.

While summative lagging indicators are important, improvement research also needs data about specific program processes and student experiences as these occur in real time. [29] This evidence is key for informing more micro-level activities linked to longer-term student success. For example, extant research suggests that the nature of students' initial engagement with their community college during the first two or three weeks of enrollment is critical.[30] Data about students' academic behaviors and experiences during these critical weeks are key to understanding whether a pathway design is reducing early disengagement. Such data also may be used formatively to improve quick outreach efforts to students before they actually disengage.

In short, the learning-through-doing orientation of a NIC requires data systems capable of informing ongoing activity. Data collection must be embedded into, rather than added on top of, the day-to-day work of program participants. [31] Inquiry now functions as a regular organizational activity rather than being thought of as a separate one-time enterprise. Consequently, improvement research requires the negotiation of an exchange relation between the time required for data collection and the utility of the information generated. To be sure, traditional psychometric concerns found in academic research still matter; but measurement in a continuous improvement context also places primacy on its informative quality for use in practice.[32] Such data must have prescriptive value, i.e., provide evidence that might help clinicians think and act better given some specific problem at hand, and be accessible in a timely fashion to inform such

---

[29] We note that these also create a basis for more micro-level process targets. In so doing, a network may catalyze the formation of sub-networks working on improving the same micro-processes and aspiring to the same common micro-targets. The overall logic of the NIC still applies but now at a more micro-level.

[30] See the extensive work on this topic using the Survey of Entering Student Engagement (http://www.ccsse.org/sense/).

[31] This idea has been developed in some detail at IHI. See: http://www.ihi.org/IHI/Topics/Improvement/ImprovementMethods/Measures/

[32] By way of example, there is great interest today in teacher assessments. Considerable attention now is directed toward developing protocols for rating classroom instruction and judging the quality of these protocols to the extent that they correlate with classroom level value-added measures of student learning. Predictive validity is viewed as the main criterion for judging instrument quality. One can envision instruments that rate relatively high by this standard, but afford little guidance as to what teachers need to learn or do differently to actually effect improvements in student learning. The latter is the Informative quality of the assessment—does it signal what we value/want others to actually attend to?

decision-making. We call this *practical measurement* and view it as a core agent structuring inquiry in a networked improvement community.[33]

UNDERSTANDING VARIABILITY IN PERFORMANCE

Closely related to the emphasis on practical measurement is a second key feature structuring empirical activity in a NIC: attention to variability in performance and the multiple factors that may contribute to it. Most field trials formally assume that there is some fixed treatment effect (also known as a standardized effect size) to be estimated. If pressed, investigators acknowledge that the estimate is actually an average effect over some typically non-randomly selected sample of participants and contexts. Given the well-documented experiences that most educational interventions can be shown to work in some places and not others, we would argue that a more realistic starting assumption is that interventions will have variable effects and these variable effects may have predictable causes. We expect, for example, that Statway effects will vary depending on specific characteristics of students, faculty and the contexts in which they both work. This perspective leads to a very different organizing question for study. Rather than asking whether an "intervention works," a network improvement community asks, "what works, when, for whom and under what sets of circumstances?"

Put somewhat differently, improvement research focuses our attention on the information necessary to make interventions work reliably at scale. 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, for different kinds of students, and for use by diverse faculty.

A COMMITMENT TO CONTRASTS AND COMPARISONS

Understanding what works when, for whom and in which contexts, also places demands on how network participants design their individual inquiries so that practical inferences can be drawn about outcome variability. In principle, we need information from each improvement cycle on the outcomes that occurred, and how these link to specific characteristics of participants, contexts and possibly time. Accumulating this evidence, and making comparisons and contrasts across it, provides the basis for examining variability both locally and across the network. It enhances possibilities for C-level learning to occur.[34]

---

[33] This is closely related to the idea of unobtrusive measures described by Webb et al. 1966.

[34] Almost four decades ago, Light and Smith (1971) detailed such an accumulating evidence strategy. While these proved formative ideas for the emergence of meta-analysis (i.e., the quantitative synthesis of research findings), Light and Smith actually cast their arguments in terms of the prospective design of a program of applied research rather than post hoc search for patterns in previously published results. It is this idea that we return to here.

Here, too, an effective compromise must be sought. Clearly, only a small number of questions can be examined at any given time and in any one place. As noted earlier, careful specification of the improvement target helps to discipline these inquiries. Tools such as the driver diagram and program improvement map assist in priority setting within this shared problem terrain. In a complementary fashion, a common inquiry protocol—the PDSA cycle—assists as well.

A PROMISING TOOL TO STRUCTURE INQUIRY: THE PDSA CYCLE

The plan-do-study-act (PDSA) cycle is a broadly used tool in improvement research across different fields (Langley et al. 1996). Used across a network, it allows activity to occur simultaneously in different contexts, but in ways that evidence actually can accumulate.[35]

The protocol below vitalizes four core questions guiding improvement research:

- How do we understand the presenting problem, including the organizational systems in which it is embedded?
- What precisely are we trying to accomplish (meaning what are the targets for the improvement research)?
- What changes might we introduce toward these ends?
- How will we know if these changes are an improvement?

We sketch below how PDSA cycles can structure disciplined inquiries by individual participants and also function as the warp and weft of a networked improvement community (see Figure 3).

---

[35] See Shavelson and Towne (2002) on the role of common methods as part of a practice of disciplined inquiry.
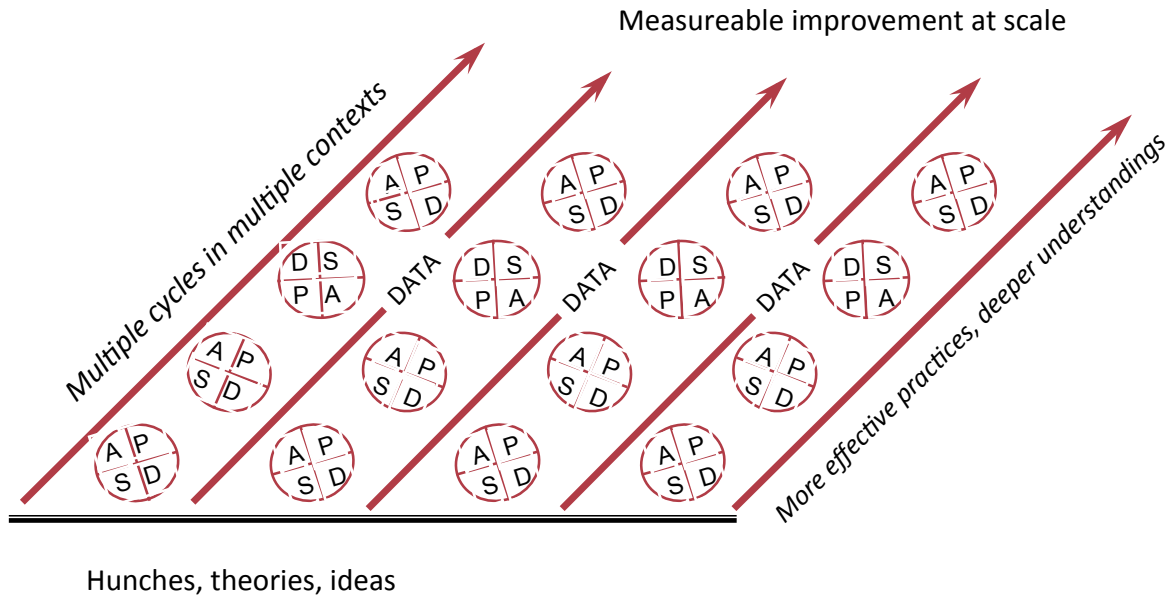
Figure 3. An Improvement Cycle
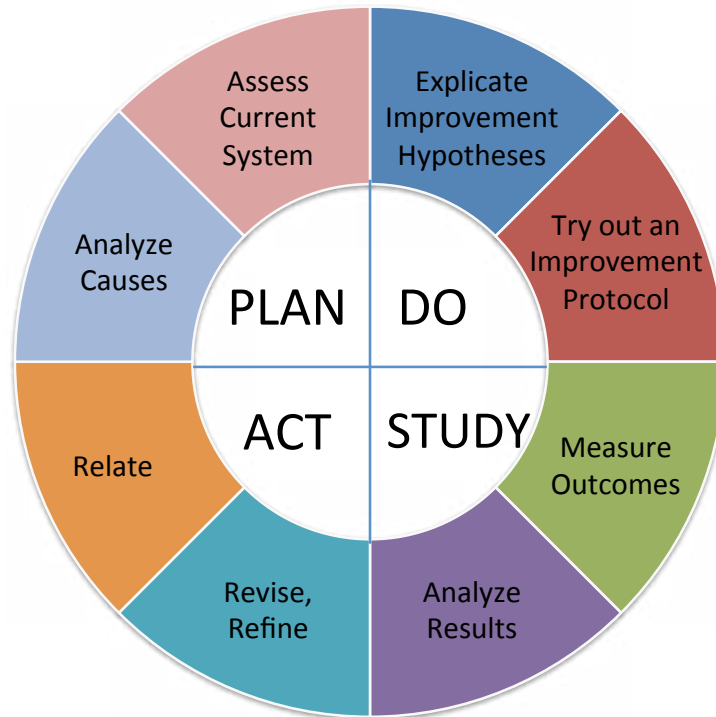
Plan. This phase directly addresses the first of the improvement research questions: How do we understand the presenting problem and the organizational system in which it is embedded? Whenever an improvement problem comes into view, educators grasp for possible solutions. They focus in on options that seem plausible given their particular professional background, organizational role and the standard operating procedures and norms of their respective organizations. Given that problems like the high failure rates in developmental mathematics are often complex system failures, these individual "point-of -view analyses" often come up short. To be sure, deep insights can be gained by viewing a problem deeply through a particular perspective. (By way of example, think of the microscope.) However, going deep also can blur our vision about the context that immediately surrounds the deep view and interact with it. (By analogy here, think of a wide-angle lens that locates a set of microscopes within a larger terrain.)

It is here where a networked improvement community benefits from shared tools such as program improvement maps and driver diagrams. Working with common frameworks during the planning phase encourages participants to build on and further explicate shared understandings of what otherwise might be tacit and partial explanations about the nature of a problem and the

larger system in which it is embedded. The planning process creates a mechanism for participants to identify and articulate locally specific knowledge and how it fits into a larger tapestry. In so doing, it structures communicative processes, anchored now in the common language system conveyed in maps and driver diagrams, that enhance network capacity to co-learn from diverse initiatives. We note that a common language framework functioned as a core organizing element in both the International Technological Semiconductor roadmap and the Linux development networks. More generally, it also has been identified as an essential characteristic of effective design communities (Norman 1988).

In addition, disciplined planning makes manifest a network narrative that all participants are researchers about practice and its improvement. As is customary in research communities, participants theorize about alternative mechanisms, plausible causes and effects. Consequently, an explicit goal for the community is to develop a working theory of practice and its improvement. Such theory likely will entail an eclectic mix of extant practices, hunches about effective new interventions, and more basic research findings. Presumably the working theory will be underspecified in the early stages of a community's work. The expectation is not perfection in its initial manifestation, but rather a good starting point. Through multiple PDSA cycles over time and contexts, the network advances on two meta goals: 1) specifying, refining, testing and accumulating more effective practices; and 2) simultaneously evolving better improvement theory to guide subsequent rounds of work. This developmental dynamic is represented in Figure 4.

Figure 4. The Learning Dynamics of a Networked Improvement Community

Measureable improvement at scale

Multiple cycles in multiple contexts

More effective practices, deeper understandings

Hunches, theories, ideas

Do. In the "Do Phase," rapid trials are launched that generate evidence about both the specific practices being attempted and the improvement hypotheses that undergird them. This phase addresses improvement questions two and three, "What specifically do we hope to accomplish and what changes will we introduce toward this end?"

The spirit of improvement research is to get a trial quickly into the field to test improvement hypotheses. Rather than trying to solve all of the institutional problems that might need to be addressed if this proposed change were to be taken up broadly (and the endless meetings that this would likely entail), network participants embrace a spirit of rapid prototyping—try it quickly, learn from it cheaply, revise and retry.[36] As corollary, in our Statway work we expect, and in fact value, that each college may implement the improvement solution somewhat differently, given local system constraints. So long as these planned differences are documented and useable data about local efficacy is gathered, the network has an opportunity to build general knowledge about whether and how an intervention can be made to work under varied circumstances.

---

[36] Since the Statway network begins as an innovation zone, this is the alpha development phase discussed in Bryk and Gomez (2008). To function as an innovation network has implications for selection of the initial charter members of the network, placing a premium on individuals and contexts conducive to such work.

Study. It is here where the fourth improvement question is directly addressed—"How will we know if the proposed change is actually an improvement?" It is human nature to believe in the efficacy of one's work, and the field of education is replete with individual testimonials about effective programs. Improvement research, however, requires adherence to rudiments of experimental design in order to create an empirical warrant for such assertions. This is captured succinctly in the phrase (often attributed to Deming): "In God we trust; all else bring data." Each PDSA cycle must establish a plausible counterfactual and test local outcomes against it. In practice, improvement researchers often employ an interrupted time series design. An outcomes baseline is established, and subsequent performance is tracked against this baseline. Observed gains over and above the baseline provide evidence of an intervention's effect. In this design, the baseline functions as the counterfactual—the outcomes we would have expected to occur absent the intervention.

As with measurement, the choice of design in improvement research is a pragmatic affair. Emphasis is afforded to nimbleness (i.e., how can we learn quickly from an individual PDSA cycle?) and practicality (i.e., how can the design of such inquiries be embedded naturally into the work life of the organization?). It is important to recognize that sturdy knowledge can accrue from a fleet of such studies, even with relatively simple designs for individual cycles of inquiry.[37] The latter especially is true when the goal is documenting large effects, and evidence of such effects can be found across multiple cycles over time and contexts. Under these circumstances, the likelihood of drawing a false generalization about improvement is greatly diminished.

In addition, it is important to remember that improvement research expects effects to vary as a function of student, staff and organizational context characteristics. Especially when the number of concomitant factors is large and unknown, we need to rely on systematic analyses of naturally occurring variation across a network to learn about the conditions under which such variation arise.[38] A richly documented fleet of studies is an essential resource in this regard.

---

[37] To be sure, randomized trials remain the strongest design to implement in improvement research when practical. However it is important to note that the results of randomized control are not always definitive. Weisberg (2010) documents that clinical trials actually can lead to bias conclusions when the causal effect of an intervention varies across cases (p.23; also Weisberg, Hayden, and Pontes 2009). Not only the magnitude, but also the direction of effects may be erroneous. Since improvement research begins with an assumption of variable effects, this caution is noteworthy.

[38] For a partial example of this see Bryk et al. (2010). Under a radical school decentralization in Chicago, significant new resources and authorities were devolved to individual schools precipitating a natural experiment in school change and improvement. Through systematic longitudinal inquiry, the authors developed in conjunction with local leaders a working theory of school improvement, a practical measurement system to characterize changes school by school, over time, and linked this in turn to a series of value-added estimates over time, to changes in student learning. Both the working theory of practice improvement that evolved here, as well as specific empirical evidence, was taken up in continuous improvement efforts in local schools and systemwide.

Act. As noted earlier, a NIC is organized around fast iterative cycles of design, engineering and development. The idea is to test fast, fail fast and early, learn and improve. Consequently, revision and refinement characterize the act phase. This phase also returns attention to systems thinking. While innovation development may focus on the design of some specific new tool, instructional resource, organizational role or routine, the act phase may raise new questions about how these artifacts interconnect with extant practices and local context. Getting these interconnections right can be key to achieving efficacy at scale.[39]

Looking at this from Englebart's C- level perspective, the network aims to develop high reliability interventions consisting of good materials, technology tools, well-specified routines, support services and so on. Rather than conceiving of scaling, solely as a matter of implementing these artifacts as designed (or what some describe as "with fidelity"), the NIC also focuses on *integrative adaptivity* as a core design problem. It assumes that any new intervention subsequently will be picked up by different participants who must make it work within their particular organizational context. Therefore, C-level activity focuses explicitly on how an innovation can be made to function well in the hands of diverse individuals working under highly varied circumstances.[40] To the point, it is not good enough to know that Statway can be made to work in a few places. The network aims to build useable knowledge for the larger field.

This is another place where a fleet of studies conducted across a network of inquiry is a special resource. In contrast to a more traditional educational R&D center, a NIC opens up possibilities for harvesting the wisdom of crowds.[41] Iterative cycles occurring within each individual site naturally focus attention on how to make the intervention work in that site. Parallel activity occurring simultaneously across a network of sites creates a naturally evolving evidence base for refining designs and generalizing how an intervention can be made to work more broadly. Within each individual site, the specificity of a local context interacts with principled design of

---

[39] The conduct of improvement research documented in the *Checklist Manifesto* provides a concrete example of this. Once Gawande (2010) had established the efficacy of the checklist in his own surgical theater, the team undertook a field study that deliberately introduced the checklist into a highly varied set of healthcare settings in terms of fiscal resources, cultural norms organizing relations among physicians and nurses, and basic organizational capacities. A key design concern at this point was whether and how this routine could be integrated into practice in organizations that were quite different than the context of original development. This is a textbook case of the problem of integrative adaptivity.

[40] Of note, both of the polar positions laid out earlier (translation research and action research) deflect attention away from this question. Under the translation paradigm, the aim is to standardize the treatment, and evidence on treatment variability is considered implementation failure. The responsibility for the latter is externalized to the local context. In action research, all of the complexity and dynamism of the context is embraced, but how an innovation might effectively travel to another locale is not generally a core subject of inquiry. In contrast, this is a core inquiry goal for a NIC.

[41] See Surowiecki 2004.

the intervention as just described.[42] Working through these transactions across a network of sites places the question of integrative adaptivity, (i.e., how do I make this work in different contexts?) squarely at the center of network-level inquiry. Such learning is key to achieving efficacy and reliably at scale. It is the journey of transversing Cornfield and Tukey's dual span from data to inference.[43]

NEXT BIG QUESTIONS

We argued in our earlier essay that educational R&D has little capacity to focus on sustained and coordinated educational problem solving. Improvement efforts abound in schools, colleges and classrooms. Academic research grows at an accelerating pace and a large market place exists for commercial goods and services. But collectively, this is not adding up in ways that advance substantial improvements at scale. As an alternative, we have introduced the idea of a networked improvement community and detailed a conceptual framework for organizing the basic rules of a new approach. Next we address the question of how such a network actually comes into existence.

EVANGELIZING LEADERSHIP

Counter to some prevailing myths, networks engaged in collective complex product development are not self-organizing (Weber 2004). In each of the effective networks we have examined, a small number of opinion leaders played a critical role in building followership and securing moral authority for organizing the rules of the game. Each, with their own style, evangelized the vision, set goals for the collective project, persuaded others of its viability and invited participation. For example, Berwick and colleagues crusaded the central tenets of quality improvement with an initially skeptical healthcare profession. They reframed medical complications as errors and provided hospitals with proof-cases that these could indeed be avoided. IHI was formed as an integrative context where healthcare professionals and institutions could come together to pursue this vision. Similarly, Torvalds recognized a niche for interested

---

[42] For an illustrative example, see how in the *Checklist Manifesto*, Gawande and colleagues systematically addressed utility of their prototype checklist by deliberately moving the checklist out to eight very different contexts. The key learning objective in this phase of the work (what we have termed beta phase inquiry) is, whether this could be made to work in very different institutional and cultural context, and if so, what would it take. This is explicit inquiry about integrative adaptivity.

[43] We note that this basic phenomenon continues in the beta phase where innovations move into new contexts. Inevitably some accommodations may be needed to integrate the initiative into these new settings. Accomplishing this well entails an analytic practice where local conditions must intersect with the principled design of the intervention (Coburn and Stein 2010). The knowledge generated at the network level, by synthesizing learning efforts at multiple sites is key to discerning how, and the conditions necessary for the intervention to be reliably engaged in other places.

programmers primed to work on an open-source basis and offered an early working version of a kernel operating system as catalyst for development of the Linux community. Likewise, a cadre of Silicon Valley leaders including Noyce, Moore, Galvin and Sporck (Schaller 2004:549) took up the mantle within their industry, arguing that cooperative efforts on a common roadmap constituted a valuable collective good in what otherwise operates as a highly competitive business environment.[44]

AN INTEGRATIVE HUB

In tandem with the evangelizing described above, initiating leaders also took on the role of designer as each created a hub for the network. As Goldsmith and Eggers (2004) detail: *"The job of this network designer is to identify possible partners, bring all the relevant stakeholders to the table, analyze current in-house operations, determine and communicate to all the members the expectations of how the network will function, assemble and enmesh all the pieces of the network, devise strategies to maintain the network and finally, activate it"* (2004:55). While living on the plane of ideas as in our earlier essay (Bryk and Gomez 2008), these considerations could easily remain unaddressed. As we sought to move these ideas into action, however, this hub function suddenly loomed very large.

In general, the hub's role is to function as an initiator of activity and an integrative force for the overall enterprise. When we looked across our three illustrative cases, a set of common objectives emerged: The hub aims to build field consensus on the importance of the problem and promising pathways to solutions. It seeks to catalyze network engagement, bringing more leaders and champions to the movement. It develops the initial version of the structuring agents and norms for participation. It maintains a technology core, such as a dynamic knowledge repository organized around the program improvement map and community use platform. It also provides analytic capacity to support B-level activity out in individual sites and has lead responsibilities for cross-institutional, C-level learning. Finally it needs to secure lines of support that flow to network participants for initiation and growth.

MANAGING MICRO-MACRO DYNAMICS

Forming such a network also requires consideration of incentives and governance. Weber (2004) notes that the effective functioning of an intentional network requires a solution to two fundamental problems in formal organization. First at the micro-level, we need to understand why people might voluntarily allocate time and attention to such networks absent normal mechanisms of compensation. Second, at the macro-level, we also must attend to how

---

[44] This evangelizing role is now being pursued by the Statway program senior partners as they reach out to community colleges, professional associations, policy and foundation leaders and the academic research community. Institutionally, the Carnegie Foundation seeks to draw on its reputation as a neutral broker and convener, as a resource in forming the connective tissue necessary for the Statway network to take root and grow.

individuals' efforts are coordinated and sustained on developing a complex product. Absent the normal mechanisms of markets and/or hierarchical bureaucratic control, a new form of network governance must be articulated. In the early stages of operation, the hub engages in the preemptive design by establishing the initial rules of the game in accordance with the key structuring agents discussed above. Over time, however, all details are open to change and change should be expected as the network grows and evolves.

CREATING INCENTIVES

Absent a capacity to direct individuals' work through compensation, a networked improvement community must depend on alternate mechanisms for incenting participants to voluntarily allocate their time and resources to a collective project. Following Weber (2004), we argue that participation in networks is not purely altruistic. Rather, participation offers individuals many non-monetary benefits documented in the network examples outlined here.

First, participants' ability to choose a specific micro-problem to work on provides opportunities for diverse individuals to deploy their particular creative energies. Within a network, the ideas generated and the artifacts produced have a natural community of appreciation, and individuals are recognized for these contributions. The Linux community, for example, provides a forum where elegant programming solutions can be shared and are acknowledged. Contributors build a reputation within the community that is recognized on a much larger stage. Important hedonic rewards are triggered through social affiliation and the according of status. At a more instrumental level, talent that may have hidden in the workplace is brought into more public view and this increases opportunities for individual professional mobility.

Second, joining a network provides access to expertise of other participants, and this enables individuals to learn new skills. In many cases, network participation can be more efficient than going about solving local problems alone. Linux programmers, for example, report that their pursuit of network tasks actually facilitates work on their primary jobs.

Third, vibrant networks also tend to propel a shared identity among participants, anchored in a common narrative of an enlightened purpose or a common enemy. In the semiconductor industry, cooperation was catalyzed by a fear that Japan would overtake United States dominance in the industry. Members of the Linux community believed that information should be free and so they banded together against their perceived enemy, which was Microsoft.

Enlivening networked improvement communities in education will require similar attention. The work structures and shared norms that we have described as fundamental to a networked improvement community depart in significant ways from the state of play in education today. Whereas engineers and software programmers are preconditioned to avoid solving the same problem twice, educators too often assume that their solutions must be invented anew in each

context.[45] This means that initiating networked improvement communities will require explicit attention to incenting different kinds of thinking and behavior and to forming new norms.

Responding to these challenges entails consideration of multiple mutually reinforcing mechanisms that extend the message that cooperative participation is highly valued. The Statway network views all participants as researchers and developers. It does not reserve this status distinction only for academics from research universities. The network aims to create numerous and varied contexts where individuals' distinct expertise can come to the fore and be used and appreciated by peers. The Carnegie Foundation will bring visibility to these individuals' contributions through its print and digital media. The network needs new leaders and champions for its future growth. This means new work arenas and possible new career opportunities for at least some participants. We will use a longstanding Foundation initiative, The Carnegie Scholars of Teaching and Learning, as a formal designation to acknowledge individuals who make major contributions and assume leadership roles. A sabbatical program offering a possible year in-residence at the Foundation represents still another mechanism for recognition. Likewise, the network seeks to incent institutional participation as well. It is important that community colleges be recognized and accorded status for their leadership. A possible elective Carnegie Classification might be used for this purpose.[46]

EVOLVING GOVERNANCE

At present we are focused on network initiation. Assuming participation is engaged, the ability of the network to sustain progress over time will depend on a crescent governance structure. In these forming days, the network hub is establishing a first iteration of the structuring agents. While the network is small, initiating leaders can serve as the main moderators of the community and in so doing establish norms for participation. As the network grows, so too does the time and energy required to make thoughtful decisions and justify them to the community. New structures must emerge to maintain collective agreements and sustain coherent future actions. In the cases we studied, network governance did not proceed along a preconceived path, but rather evolved over time in response to needs and conflicts that arose in the process of joint work. We expect that to happen in Statway as well. Over time, initiating structures likely will require multiple iterations of refinement and possible larger changes to accommodate network growth and movement toward becoming self-sustaining. We anticipate that tensions will need to be negotiated and accommodations made. While the final shape and organization of the network

---

[45] It is interesting to note that these same arguments appear in the early history of the quality improvement movement in health services. Kenney's (2008) account details exchanges in this regard. There are several places in his text where one could easily exchange the words "doctors and hospitals" with "teachers and schools."

[46] There is a precedent in the elective community engagement classification that Carnegie established in 2005. Participation is voluntary and over 195 colleges have chosen to do so. It involves a detailed, data-based process of application and membership which has proven quite meaningful across the larger community.

remains emergent, the clear intent is for the network governance to broaden beyond the initiating hub.

CHOOSING A LICENSE

In an open resource world, the license explicates the rights and responsibilities of network membership and creates ground rules for how pieces of the work are shared. It structures social transactions around intellectual property. Early in the development of Linux, Torvalds made the somewhat controversial decision to distribute the kernel code under a General Public License. At that point in the community's evolution, Torvalds single-handedly made the major decisions, but he did so with careful attention to what the community wanted, because product production depended on attracting large numbers of developers to work on the code. His license decisions were not based on any commitment to this particular type of license per se. Rather, the license was an instrument to facilitate the work of the community and scale participation.[47]

Clearly, the choice of an appropriate license is key for growth of a networked R&D community. We are committed in principle to open resources and believe that all intellectual property derived by the network belongs to the network. The Foundation's role is to act as steward in this regard. The precise form of a license to deploy, however, remains an open question. We know that the license must incent the contributions of individual practitioners, researchers and educational designers who operate in this space, but also those who bring different interests and seek different benefits from network participation. In short, a license must fit into a networked improvement community, not the other way around. A best practice is one that grows and sustains participation, focuses ongoing efforts on targeted priorities, and ultimately contributes to improvement reliably at scale.

THE WORK AHEAD

Statway is Carnegie's first attempt to take on the integrative role of a network hub. We are working off of a set of empirically grounded hypotheses about how best to initiate and integrate a network that aims at social learning and complex product development. Through doing this, Carnegie will learn much about the essential functions of a network integrator.

This chapter represents our evolving thinking to date on these matters. Our hub's first efforts have focused on creating vital connections out to colleges that afford a powerful context for innovation co-development. Drawing commercial partners into this work and thinking about how mutual benefit partnerships might be best structured here remains work on the horizon. Likewise,

---

[47] See Weber (2004), pg. 111-116.

expanding engagement of the applied research community lies ahead. We have some forming ideas for initiating outreach in both of these domains, but that will be pursued in the future.

The Statway network is in an alpha or initiating phase. Its priorities now are concrete and practical. Can the hub co-develop with community colleges promising pathways for student success? Can it catalyze, sustain and grow social participation in the charter colleges and beyond? Will it generate sufficient enthusiasm among the faculty in the initial network (including a sense of efficacy in their teaching and encouraging initial evidence on student learning) so that there is an eagerness to persist and recruit other colleagues into the work? Most important, can it vitalize both B-level and C-level learning for improvement? These are the primary concerns for evaluating alpha development in a networked improvement community.

Engaging disciplined-based research and researchers is essential throughout this work. Can emerging principles in cognitive science, for example, be translated into scalable instructional practices within Statway? Similarly, concerns about student engagement and motivation play a key role in the overall problem system. Can findings in social psychology, for example on identity development and stereotype threat, be exploited and in the process further tested here? Language and literacy issues also abound. Can researchers from these domains study and inform how the text, tasks, tests, and talk of mathematics instruction are made more productive? These are but a few of the places where disciplinary theories meet practice and ultimately are tested and refined in the cauldron of making ideas work in action.

Assuming a successful initiation of the Statway NIC, phase II will focus on network growth. While original co-development efforts will continue in the initial set of college sites, a new inquiry objective moves into primary position in the beta phase. How do we make prototype interventions function reliably at scale in the hands of a more diverse faculty and working in more varied organizational contexts? Issues of institutional change come center stage and conceptual frameworks from organizational sociology and political science are key resources. Likewise as efforts scale, we also must assemble more nimble, robust and practical design and measurement strategies to continue to learn from practice and establish increasingly deep evidence warrant for the overall enterprise. Expertise in psychometrics and sociometrics becomes essential.

Subsequently, improvement becomes institutionalized in phase III. Core guidance shifts from the hub into the network that has evolved into a self-governed, professional community of practice improvement. Operating now at a very large scale, the network continues to invest in data mining strategies and other mechanisms as it seeks to continue to learn and improve from an ever-enlarging base of network level action. This is educational statistics and data analysis at large scale.

As we focus on problems of phase I network initiation, we also are attending to this developmental arc. We aim to supply a base of practical and robust artifacts, processes, work roles and routines coupled with evidence of costs and efficacy that can be adaptively integrated by others. Likewise, we already are considering alternative strategies for the human and social resource development that will be needed in the future by others who wish to engage Statway in their particular circumstances. Further, we are moving to expand the network of academic expertise necessary to inform these developments.

Simultaneously, we also are advancing the demand side by focusing on both top down strategies that aim to engage policy advocates and grass roots strategies to mobilize faculty, staff and community college leaders. From the start we began developing for scale and see this task as both creating a supply innovation and catalyzing a demand for change. In conceptualizing scaling as a problem of collective learning, we target both supportive policy actions and seek to engage the minds and hearts of community college leaders, faculty and staff. Ultimately, this is where scaling with efficacy will be either succeed or fail in classrooms, schools and colleges where students and educators must join together to advance learning.

In closing, we note that we have used the example of Statway to illustrate the themes of this chapter. We wish to reiterate that Statway is just one attempt to redress a larger concern—developing a more effective educational R&D infrastructure. A field is now emerging around new approaches to applied inquiry on problems of practice improvement. Each effort will entail multiple cycles of institutional design, engineering and development. These too will follow a developmental arc, and structuring opportunities for community learning can accelerate improvement here, too. Much knowledge know-how can be gleaned from comparative analyses across multiple cases emerging in the field. Our hope is that this chapter functions as one such convening context where diverse individuals, sharing the common concern of strengthening educational R&D—might join in analysis and critique, and ultimately advance a next stage in the evolution of educational improvement.

REFERENCES

Bailey, T., D. W. Jeong, and S. W. Cho. 2008. "Referral, Enrollment, and Completion in Developmental Education Sequences in Community Colleges." CCRC Working Paper No. 15, Teachers College, Columbia University, New York.

Berwick, D. M. 2008. "The Science of Improvement." *The Journal of the American Medical Association* 299(10):1182-1184.

Boudett, K. P., E. A. City, and R. J. Murnane, eds. 2005. *Data Wise: A Step-By-Step Guide to Using Assessment Results to Improve Teaching and Learnin*g. Cambridge: Harvard Education Press.

Bryant, M. J., K. A. Hammond, K. M. Bocian, M. F. Rettig, C. A. Miller, and R. A. Cardullo. 2008. "School Performance Will Fail to Meet Legislated Benchmarks." *Science* 321.

Bryk, A. S. 2009. "Support a Science of Performance Improvement." *Phi Delta Kappan* 90(8):597-600.

Bryk, A. S. and L. M. Gomez. 2008. "Ruminations on Reinventing an R&D Capacity for Educational Improvement." Pp. 181-206 in *The Future of Educational Entrepreneurship: Possibilities of School Reform,* edited by F. M. Hess. Cambridge: Harvard Education Press.

Bryk, A. S., P. B. Sebring, E. Allensworth, S. Luppescu and J. Q. Easton. 2010. *Organizing Schools for Improvement: Lessons from Chicago*. Chicago: University of Chicago Press.

Burkhardt, H. and A. H. Schoenfeld. 2003. "Improving Educational Research: Toward a More Useful, More Influential, and Better-Funded Enterprise." *Educational Researcher* 32(9):3-14.

Campbell, D. T. and J. C. Stanley. 1963. "Experimental and Quasi-Experimental Designs for Research and Teaching." Pp. 84 in *Handbook of Research on Teaching,* edited by N. L. Gage. Chicago: Rand McNally.

Coburn, C., E. and M. K. Stein, eds. 2010. *Research and Practice in Education: Building Alliances, Bridging the Divide*. Lanham, MD: Rowman and Littlefield Publishers.

Cook, T. D. and D. T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally College Publishing Company.

Cornfield, J. and J. W. Tukey. 1956. "Average Values of Mean Squares in Factorials." *Annals of*

Mathematical Statistics 27:907-49.

Cronbach, L. J. 1980. *Toward Reform of Program Evaluation*. San Francisco: Jossey-Bass.

Cullinane, J. and U. Treisman. 2010. "Improving Developmental Mathematics Education in Community Colleges: A Prospectus and Early Progress Report on the Statway Initiative." NCPR Working Paper, National Center for Postsecondary Research, New York.

Deming, W. E. 2000. *Out of the Crisis*. Cambridge: MIT Press.

Englebart, D. C. 1992. *Toward High-Performance Organizations: A Strategic Role for Groupware*. Groupware '92. San Jose, CA: Morgan Kaufman Publishers.

Englebart, D. C. 2003. *Improving Our Ability to Improve: A Call for Investment in a New Future*. IBM Co-Evolution Symposium.

Gawande, A. 2007. *Better: A Surgeon's Notes on Performance*. New York: Henry Holt.

Gawande, A. 2009. *Checklist Manifesto: How to Get Things Right*. New York: Metropolitan Books.

Goldsmith, S. and W. D. Eggers. 2004. *Governing By Network: The New Shape of the Public Sector*. Washington, DC: Brookings Institution Press.

Gomez, L. M., K. Gomez, and B. R. Gifford. 2010. "Educational Innovation with Technology: A New Look at Scale and Opportunity to Learn." *Educational Reform: Transforming America's Education through Innovation and Technology*. Whistler, BC: Aspen Institute Congressional Conference Program Papers.

Hiebert, J., R. Gallimore, and J. W. Stigler. 2002. "A Knowledge Base for the Teaching Profession: What Would It Look Like and How Can We Get One?" *Educational Researcher* 31(5): 3-15.

Juran, J. M. 1962. *Quality Control Handbook*. New York: McGraw-Hill.

Kelly, G. J. 2006. "Epistemology and Educational Research." Pp. 33-56 in the *Handbook of Complementary Methods in Educational Research,* edited by J. L. Green, G. Camilli, and P. Nelmore. Mahwah, NJ: Erlbaum.

Kenney, C. 2008. *The Best Practice: How the New Quality Movement is Transforming Medicine*. New York: Public Affairs.

Langley, G. J., R. D. Moen, K. M. Nolan, T. W. Nolan, C. L. Norman and L. P. Provost. 1996. *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance*. San Francisco: Jossey-Bass.

Lewin, K. 1942. "Field Theory of Learning." *Yearbook of National Social Studies of Education* 41:215-242.

Light, R. J. and P. V. Smith. 1971. "Accumulating Evidence: Procedures for Resolving Contradictions among Different Research Studies." *Harvard Educational Review* 41:429-471.

National Academy of Education. 1999. "Recommendations Regarding Research Priorities: An Advisory Report to the National Education Research Policy and Priorities Board." New York: NAE.

Norman, D. 1988. *The Design of Everyday Things*. New York: Currency Doubleday.

Paley, A.R. 2007,. "'No Child' Target is Called out of Reach." *The Washington Post,* March 14. Retrieved September 1, 2010 (http://www.washingtonpost.com/wp-dyn/content/article/2007/03/13/AR2007031301781.html).

Podolny, J. M. and K. L. Page. 1998. "Network Forms of Organization." *Annual Review of Sociology* 24:54-76.

Powell, W. W. 1990. "Neither Market nor Hierarchy: Network Forms of Organization." *Research in Organizational Behavior* 12:295-336.

Ruopp, R. G., S. Gal, S. Drayton and B. Pfister 1992 *Labnet: Toward a Community of Pactice*. Hillsdale: Lawrence Erlbaum Associates.

Schaller, R. R. 2004. "Technological Innovation in the Semiconductor Industry: A Case Study of the International Technology Roadmap for Semiconductors (ITRS)." *School of Public Policy*. Fairfax, VA: George Mason University.

Schlager, M., J. Fusco, and P. Schank. 2002. "Evolution of an On-Line Education Community of Practice." *Building Virtual Communities: Learning and Change in Cyberspace*. K. A. Renninger and W. Shumar. NY: Cambridge University Press.

Shavelson, R. J. and L. Towne, eds. 2002. *Scientific Research in Education*. Washington DC: National Academy Press.

Shirky, C. 2008. *The Power of Organizing without Organizations*. New York: Penguin Press.

Committee on a Strategic Education Research Partnership (SERP). 2003. Washington DC: Strategic Education Research Partnership.

Surowiecki, J. 2004. *The Wisdom of Crowds*. New York: Anchor Books.

von Hippel, E. 2005. *Democratizing Innovation*. Cambridge: The MIT Press.

Webb, E. J., D. T. Campbell, R. D. Schwartz and L. Sechrest. 1966. *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago: Rand McNally.

Weber, S. 2004. *The Success of Open Source*. Cambridge, MA: Harvard University Press.

Weisberg, H. I. 2010. *Bias and Causation: Models and Judgment for Valid Comparisons*. Hobokon, NJ: John Wiley and Sons.

Weisberg, H. I., V. C. Hayden, and V. P. Pontes. 2009. "Selection Criteria and Generalizability within the Counterfactual Framework: Explaining the Paradox of Antidepressant-Induced Suicide." *Clinical Trials* 6(2):109-118.

Wenger, E. 1999. *Communities of Practice: Learning, Meaning and Identity*. Cambridge: Cambridge University Press.