Large Scale High School Reform through School Improvement Networks: Examining Possibilities for "Developmental Evaluation"

Donald J. Peurach
University of Michigan

Sarah Winchell Lenhoff
Michigan State University

Joshua L. Glazer
The Rothschild Foundation

Paper presented at the 2012 Conference of the
National Center on Scaling Up Effective Schools
Nashville, TN: June 10-12, 2012

Author Note

Donald J. Peurach, School of Education, University of Michigan; Sarah Winchell Lenhoff, Michigan State University; Joshua L. Glazer, The Rothschild Foundation.

The authors gratefully acknowledge funding received from the School of Education at the University of Michigan and from the Education Policy Center at Michigan State University.

Address all correspondence to: Donald J. Peurach, School of Education, University of Michigan, 610 E. University, Ann Arbor, MI, 48109. E-mail: dpeurach@umich.edu.

ABSTRACT

The following analysis has two aims: to examine the potentially negative consequences of summative impact evaluations on school improvement networks as a strategy for large scale high school reform; and to examine formative "developmental evaluations" as an alternative. The analysis suggests that it is possible to leverage theory and research to propose meaningful criteria for developmental evaluation, and a developmental evaluation of a leading, high school-level school improvement networks suggests that these criteria are useful for generating formative feedback for network stakeholders. With that, the analysis suggests next steps in refining formal methods for developmental evaluation.

Keywords: evaluation, impact evaluation, developmental evaluation, best practice, educational reform, innovation, knowledge production, organizational learning, replication, scale, school turnaround, sustainability

Large Scale High School Reform through School Improvement Networks: Examining Possibilities for "Developmental Evaluation"

The national education reform agenda has rapidly evolved to include a keen focus on large-scale high school improvement. In contrast to targeted interventions, one promising reform strategy is school improvement networks in which a central, "hub" organization collaborates with "outlet" schools to enact school-wide designs for improvement: for example, as supported by comprehensive school reform providers, charter management organizations, and education management organizations (Glazer and Peurach, 2012; Peurach and Glazer, 2012).[1] Examples include the Knowledge is Power Program, the New Tech Network, and Green Dot Public Schools.

Over the past twenty years, school improvement networks have benefitted from billions of dollars in public and philanthropic support, largely on their perceived potential to support rapid, large-scale improvement in student achievement. Even so, research on the management, implementation, and effectiveness of comprehensive school reform programs suggests that school improvement networks emerge and mature over very long periods of time -- decades, in some cases (Berends, Bodilly, & Kirby, 2002; Borman, Hewes, Overman, & Brown, 2003; Glennan, Bodilly, Galegher, & Kerr, 2004; Peurach, 2011). Further, research suggests that their emergence and maturation is highly dependent on coordinated environmental supports (Glazer and Peurach, 2012), with federal policy a key component and driver of such supports (Bulkley and Burch, 2012; Peurach, 2011).

[1] We distinguish school improvement networks from the "networked improvement communities" advanced by the Carnegie Foundation for the Advancement of Teaching (Bryk, Gomez, & Grunow, 2010). In school improvement networks as defined here, the hub functions as the primary locus of design and as the chief agent supporting the replication of "best practices" across outlets. More in keeping with ideas of "open source" networks, the hub in networked improvement communities establishes an infrastructure to support (and insure the integrity of) distributed design and problem solving activity among outlets. For more on this comparison, see Clyburn (2011).

The disconnect between expectations for rapid success and slow rates of emergence and maturation can leave school improvement networks vulnerable to rapid shifts in environmental support, both individually and en masse. For example, in the case of comprehensive school reform, the failure of all but a small number of programs to quickly provide rigorous evidence of positive, significant, and replicable effects on student achievement was instrumental in the rapid dissolution of environmental supports and the subsequent decline of the movement, despite billions of dollars in public and private investment and despite the potential loss of formidable intellectual capital in failed networks (Peurach and Glazer, 2012).

For those who see potential in school improvement networks, the preceding suggests a need to stabilize the agenda for high school improvement to create the time required for networks operating at the high school level to emerge and mature. That, in turn, requires complementing conventional impact evaluations with new "developmental evaluations." Conventional impact evaluations are largely summative in nature, and designed to identify the replicable effectiveness of school-wide improvement programs. By contrast, new developmental evaluations would be formative in nature, and designed to provide evidence of strengths and vulnerabilities that have potential to support (or to undermine) replicable effectiveness.[2]

Developmental evaluations would be useful to policy makers, philanthropists, and other decision makers to assess progress and to guide funding decisions. They would be useful to practicing reformers in improving the structure and function of school improvement networks.

[2] As discussed in this paper, our notion of developmental evaluation is not entirely consistent with that advanced by Patton (2006; 2011). Patton's approach to developmental evaluation is presented as an alternative not only to summative impact evaluation but, also, to formative evaluation en route to summative impact evaluation (which is the approach that we discuss and develop here). While we lean strongly toward Patton's approach, the place of summative impact evaluation in contemporary educational reform has us beginning our work by considering developmental evaluation in interaction with summative impact evaluation.

And they would be useful to schools and districts in selecting school improvement networks with which to partner.

The problem, however, lies in the lack criteria for assessing the development of school improvement networks. Short of statistically significant effects on student achievement, those vested in school improvement networks lack a small number of readily investigated markers that could be used to demonstrate progress, argue for agenda stability, and improve operations.

Thus, the purpose of this analysis is to propose and investigate criteria for the developmental evaluation of school improvement networks. Our argument is that it is possible to leverage theory and research to propose meaningful criteria for developmental evaluation, and our investigation suggests value in using these criteria to generate formative feedback for an array of stakeholders.

We structure our analysis in four parts. In the first part, we critically analyze the conventional evaluation paradigm, especially as rooted in assumptions that school improvement networks emerge and mature in accordance with a sequential, diffusion-oriented logic. In the second, we propose an alternative, evolutionary logic and associated criteria as the basis for developmental evaluation, anchored in an understanding of school improvement networks as contexts for collaborative, experiential learning. In the third, we demonstrate the power of these criteria by using them to structure a developmental evaluation of the New Tech Network, a leading, high school level school improvement network. In the fourth, we reflect on all of the preceding in considering possibilities for further advancing the practice of developmental evaluation.

Conventional Evaluation: Goals, Processes, and Challenges

We begin by critically analyzing conventional methods of evaluating externally developed educational improvement programs, including school improvement networks. We first examine the goals, processes, and challenges of conventional evaluation, and we conclude by discussing considerations for alternative methods of evaluation.

Goals: Identifying a Replicable Treatment Effect

Evaluations of externally developed educational improvement programs typically have two goals (Raudenbush, 2007; Slavin and Fashola, 1998). The first goal is to identify a "treatment effect" that demonstrates program impact on relevant outcomes. As a minimum standard, a treatment effect would be evidenced by a positive, statistically significant difference in achievement between students who participated in a particular program and students who did not. As a more rigorous standard, a treatment effect would be further evidenced by results establishing a causal relationship between the treatment and outcomes. The second goal is to identify whether the treatment effect can be replicated beyond early adopting school(s) and in a broader pool of schools. Cast in terms of school improvement networks, these goals result in two driving questions: (1) Is the school-wide model that functions as the foundation of the network effective in improving student achievement as compared to some counterfactual? (2) Can program effects be replicated in newly-adopting schools?

The pursuit of replicable treatment effects, in turn, is linked tightly to common conceptions of "scaling up" externally-sponsored educational improvement initiatives. For example, Schneider and McDonald (2007a:4) define scale up as "the enactment of interventions whose efficacy has already been established in new contexts with the goal of producing similarly positive impacts in large, frequently more diverse populations." Summarizing alternative conceptions, Constas and Brown (2007:253) define scale up as "the process of

testing the broad effectiveness of an already-proven educational intervention as it is implemented in large numbers of complex educational contexts."

Over the past ten years, providers of externally developed educational improvement programs (including those sponsoring school improvement networks) have faced increasing pressure to provide rigorous evidence of replicable treatment effects. This pressure derives from multiple sources: for example, the broader standards-and-accountability movement in education; criteria that link program adoption and continued funding to rigorous evidence of replicable effectiveness; the founding of the Institute of Education Sciences in 2002, and its mission to identify "what works, what doesn't, and why" (Institute for Education Sciences, 2012a); and the emergence of organizations such as the What Works Clearinghouse and the Best Evidence Encyclopedia, which link the legitimacy of programs to rigorous evidence of replicable effectiveness. Concern with the replicable effectiveness of educational programs mirrors efforts to establish the impact of other social programs in the US and abroad (Granger, 2011; Khandker, Koolwal, and Samad, 2010).

Process: A Four Stage Evaluation Process

Efforts to use impact evaluations to establish the replicable effectiveness of externally developed programs (including school improvement networks) are often organized using a four stage process, with each stage marking an increase in the number of participating schools, the standards of evidence, and, thus, the costs and sophistication of evaluation. Briefly, the stages are as follows: (1) evaluate a proposed program for its use of scientifically-based research or other sources of "best practice;" (2) implement the program in one or a small number of schools to establish "proof of concept," with success evidenced via descriptive and other qualitative studies; (3) increase the installed base of schools and use more rigorous research methods (e.g.,

matched-comparison designs) to examine the magnitude and statistical significance of program effects on student achievement; and (4) further increase the installed base and use even more rigorous research methods (e.g., quasi-experimental designs, randomized control trials, and meta-analyses) to further examine the magnitude and significance of program effects.

A combination of issues (e.g., funding cycles, the need to ensure due diligence, and the desire to capitalize quickly on investments) often interact to drive the four stage evaluation process along a predictable timeline.[3] The first stage (establishing a basis in research and/or best practice) is enacted prior to implementation over a one-to-two-year window. The second stage (establishing proof of concept) is typically enacted in a one-to-three-year window. The third stage (generating evidence of effectiveness) is typically enacted in two-to-four-year window. The fourth stage (generating rigorous evidence of effectiveness while operating at a large scale) is typically enacted in a three-to-five-year window. With those as estimates, the large-scale replication of effective programs can (in principle) be accomplished in as little as seven years and as many as fourteen years.

This four-stage evaluation process is coupled closely with conventional assumptions that the development of educational interventions adheres to a sequential innovation process. This process is anchored in a diffusion-centered logic by which knowledge (in the form of basic and applied research) is put into practice at a large scale. Educational researchers have framed this model as an "RDDU" sequence: research, development, dissemination, and utilization (Rowan, Camburn, & Barnes, 2004). Others have framed this model as a stage-wise innovation process: needs/problems definition; basic and applied research; development, piloting, and validation; commercialization; and diffusion and adoption (Rogers, 1995).

[3] Time estimates are derived from Institute for Education Sciences (2012b).

This diffusion-centered logic is highly institutionalized. For example, to support the development and dissemination of research-based and research-proven school-wide improvement models, the New American Schools initiative drew directly from the sequential model of innovation to structure support as a six year, four phase progression: competition and selection, development, demonstration, and scale up (Bodilly, 1996). Currently, the Institute for Education Sciences' goals and funding criteria are consistent with this innovation process: identification projects; development projects; efficacy and replication trials; and scale up evaluations (U.S. Department of Education, 2012b). Further, the sequence of development, validation, and scale-up grants within the federal Investing in Innovation (i3) program reflects this same innovation process in supporting (among other initiatives) the development and scale up of school improvement networks (U.S. Department of Education, 2010).

Challenges: Threats to Conventional Evaluation

Though used widely in education, this conventional, four stage impact evaluation progression is vulnerable to challenges that complicate both completing evaluations and drawing valid inferences about the replicable effectiveness. Some of these challenges arise in schools: for example, the potential for program abandonment as a consequence of internal and/or external turbulence; the possibility that schools are enacting many simultaneous "treatments"; and the possibility that control schools are, themselves, enacting many simultaneous "treatments." Additional challenges arise from this mode of evaluation: for example, the problem of impact evaluation drawing dollars and attention away from program development and implementation; the lack of consensus in education on research design, criteria for incorporating studies into meta-analyses, and standards for interpreting effect sizes;

and the fact that knowledge, capabilities, and capacity for sophisticated impact evaluations are every bit as emergent as school improvement networks.[4]

Still other of these challenges are anchored in the realities and complexities of developing and scaling up school improvement networks. These challenges can be understood as arising in and among schools, programs, hub organizations, and environments (Cohen et. al, in press; Peurach, 2011).

Challenges in Schools

One challenge is that the schools that serve as the "subjects" in conventional impact evaluations are not stable entities but, instead, are entities that are fundamentally reconstituted over the course of any longitudinal evaluation. Indeed, what distinguishes school improvement networks from other large-scale reform strategies is that these networks take the entire school as the unit of treatment: not just their formal roles, structures, and technologies but, also, the teachers and leaders who comprise the school, their individual capabilities and motivations, and their collective capabilities and culture. However, all schools are vulnerable to student, teacher, and leader transiency, with chief targets of school improvement networks (underperforming schools serving large populations of at-risk students) particularly vulnerable. Consequently, the social make up of any given "subject" changes continuously, sometimes within a school year and nearly always between school years. From a social perspective, the subject in Year 1 simply is not the same subject as in Years 2 and beyond and may, in fact, be fundamentally different, and for reasons that have nothing to do with the treatment.

[4] Consider, for example, that the Society for Research on Educational Effectiveness (the chief professional organization focused on understanding cause-effect relationships in educational programs and interventions) was only established in 2005. Further, consider that issues related to the potential and problems of summative impact evaluation have been (and continue to be) hotly debated among both proponents and critics (e.g., Foray, Murnane, and Nelson, 2007; Mosteller and Boruch, 2002; Schneider and McDonald, 2007b). Finally, consider that recent emphasis on impact evaluations in other domains of social improvement have led to political, empirical, and practical challenges described as both dividing and overwhelming evaluators (Easterly, 2009; Khandker, Koolwal, and Samad, 2010).

Further, the prospects for such turbulence are exacerbated by time issues related to the enactment and evaluation of external models for school-wide improvement. To start, there is the lengthy "production function" in schools: six or seven years for elementary schools students; two to three years for middle school students; and four years for high school students. Then there is the lengthy implementation process, with the possibility of an individual school needing years to fully operationalize an externally-developed, school-wide improvement model.[5]

Finally, owing to policy pressures and incentives, school improvement networks often target schools either with no demonstrated capabilities (e.g., newly-created charter schools) or with very weak capabilities (e.g., underperforming schools). This includes capabilities either to incorporate and use external resources or to adapt and improve external resources in response to local problems and needs.[6] Yet, as the "subjects" within conventional impact evaluations, these schools are to quickly incorporate and enact exceedingly complex "treatments," and to quickly learn from experience to improve their operations. That, in turn, presents steep challenges for demonstrating replicable effectiveness on conventional impact evaluations.

Challenges in Programs

A second challenge is that the "treatment" (i.e., the school-wide improvement model) typically varies: at any point in time and over time; within and between schools; and among components of the model. This variation can be an artifact of the process by which school-wide

[5] For example, in a meta-analysis of the effects of comprehensive school reform programs on student achievement, Borman, Hewes, Overman, and Brown (2003:153) reported that achievement gains for schools that had implemented comprehensive school reform programs for five years or more were two times the overall average achievement gains, that gains for schools that had implemented programs for seven or more years were 2.5 times greater, and that schools that had implemented programs for eight to fourteen years were 3.3 times greater. Further, combining evidence across all reviewed programs, Borman and colleagues actually reported slight *decreases* in adjusted effect sizes between years 1 and 4, a finding that they hypothesized as attributable to the notion of an "implementation dip" that occurs early in the implementation of educational interventions, as teachers and school leaders "unlearn" and "relearn" their practice.

[6] Later in our analysis, we will refer to the former as "absorptive capacity" and the latter as "dynamic capabilities". See Cohen and Levinthal (1990) and Dosi, Nelson, and Winter (2001).

improvement models emerge and mature. Rather than by an RDDU-like sequence, research on networks operated by comprehensive school reform providers and charter management organizations suggests that pressures and incentives to rapidly initiate large scale operations have hub organizations attempting to replicate organizational models that are partial, problematic, and, thus, under continuous improvement (Berends, Bodilly, & Kirby, 2002; Cohen et. al, in press; Glennan, Bodilly, Galegher, & Kerr, 2004; Peurach and Glazer, 2012). A consequence is that mature models for school-wide improvement do not exist *in advance scaling up* but, instead, develop and mature *through the process of scaling up* over time. Thus, the "treatment" (i.e., the school-wide improvement model) actually changes from year-to-year as a consequence of experimentation and experiential learning within the network.

Further, variation in the treatment can be intentional, such that the causal mechanism actually varies between schools. For example, some school improvement networks intentionally delegate formidable responsibilities for design and problem resolution to schools to manage in the context of implementation. Reasons for doing so include deference to (and the desire to capitalize on) local expertise; variation in the needs of students; variation in context (school, community, district, and state); ideological commitments to local control and professional autonomy; and/or the lack of resources or capabilities to provide detailed guidance for practice. Examples of school improvement networks that feature intentional variation in the "treatment" include Accelerated Schools Plus, America's Choice, Success for All, the Knowledge is Power Program, the Big Picture Company, and the New Tech Network.

Finally, variation in the treatment can result from a paradox of school improvement networks: The treatment and the subject are confounded. While the ostensible "treatment" is an external (and, often, dynamic and adaptive) school-wide improvement model, that treatment as

enacted in schools is the product of interdependent activities among specific teachers and leaders in specific (and often weak) schools. Consequently, the "treatment" varies with their initial and developing understandings, capabilities, values, and norms; changes in the social constitution of the school (as described above); personal, organizational, district, and community histories; and much more. Indeed, borrowing from Cohen, Moffitt, and Goldin (2007), the problem (i.e., a new and/or underperforming school in need of school-wide improvement) is, itself, the solution (i.e., the mechanism by which an external design is understood, enacted, and used to improve student achievement).

One possible way to manage the problem of "treatment variation" would be to place a premium on "fidelity of implementation." While adaptive for purposes of evaluation, insistence on fidelity of implementation could be maladaptive for the school improvement network itself, in that it could actually undermine implementation and effectiveness by limiting efforts to align with local contexts and/or learn from experience. Indeed, in their research on the Big Picture Company, McDonald, Klein, & Riordan (2009) describe "the fidelity challenge" as one of eight challenges endemic to the scale up of school-wide designs: "Ignore fidelity and what will you take to scale? Ignore adaptation and your design will crack. This is more than just a challenge. It is a dilemma. It can only be managed, never resolved" (p. 19).

Challenges in Hub Organizations

A third challenge lies in the capabilities of hub organizations to develop, administer, evaluate, and refine the "treatment." That such capabilities exist is a tacit assumption of conventional impact evaluations. However, as with the federal i3 program, funding to support the development and scale up of school improvement networks is often awarded either to (a) newly-emerging hub organizations with little or no demonstrated capabilities to support large-

scale, school-wide improvement or (b) existing hub organizations that are poised to expand the breadth and scale of their operations beyond their current base of experience.

Indeed, longitudinal research on leading comprehensive school reform programs suggests that, rather than existing in advance of scaling up the network, capabilities for such work emerged through the work of scaling up the network, through a decade or more of organizational development and experiential learning (Cohen et. al, in press; Peurach, 2011). Particularly problematic was the development of large cadres of expert field staff capable of collaborating with new and/or underperforming schools to make effective use of evolving (and potentially problematic) programs. Indeed, just as the "subject" and the "treatment" are confounded, so, too, are the "treatment" and its "administrator."

Challenges in Environments

A fourth challenge is that school improvement networks operate in environments that complicate both their work and summative evaluations of their work. As compared to countries with strong coordination among national curriculum, standards, assessments, and professional education, US educational environments are argued to provide little such "educational infrastructure.". Rather, they have long been -- and still are -- characterized by emerging and uncoordinated state standards and assessments; weak professional knowledge and education for teachers and school leaders (and weaker yet for external coaches); a weak, conservative "school improvement industry" providing component technologies that support practice and its improvement; weak oversight of that school improvement industry; all compounded by incoherence, fragmentation, and turbulence (Cohen and Moffitt, 2009; Cohen et. al, in press;

Cohen and Spillane, 1991; Hess, 1999; Meyer, Scott, and Deal, 1983; National Governors Association, 2008; Rowan, 2002; Smith and O'Day, 1991).[7]

Thus, while the conventional, RDDU logic is predicated on a stock of foundational and practical knowledge to support the development of school improvement networks, there is much to suggest otherwise. Consequently, school improvement networks must compensate for environmental weaknesses by creating necessary knowledge, component technologies, and human resources. Consider the case of Success for All: a leading comprehensive school reform provider founded by researchers affiliated with Johns Hopkins University, committed to research-based educational reform, and focused on K-6 reading (a comparatively highly developed knowledge domain); yet which depended heavily on collaborative, experiential learning with schools to generate the practical knowledge and component technologies needed to demonstrate replicable effectiveness (Peurach, 2011; Peurach and Glazer, 2012).

Conspicuously absent in US educational environments is knowledge of how to organize, manage, improve, and sustain the hub organizations responsible for establishing and operating school improvement networks: a novel category of educational practice in the US, but a category about which there is little theoretical or practical knowledge and no established tradition of professional preparation (Peurach, 2012; Peurach and Gumus, 2011). Absent

[7] Consider the following critique, from the TeachingWorks initiative in the School of Education at The University of Michigan, a pioneering effort to establish a professional system supporting the practice of teaching: "After more than one hundred years of organized professional education for teachers in the United States, we still lack a clear specification of the most essential tasks and activities of classroom teaching. The curriculum for learning teaching comprises theoretical knowledge and instructional 'methods', but there is no agreement about either the knowledge that matters for teaching or what constitute effective 'methods.' Professional bodies such as the Interstate Teacher Assessment and Support Consortium (InTASC) stipulate that teachers need to know and use a 'variety of instructional strategies,' but what are these strategies? Licensure assessments for those entering teaching reflect this uncertainty; virtually all measure some aspects of candidates' personal content knowledge but few test their knowledge at a standard adequate for teaching it, and even fewer require evidence of performance ability—in part because there is no professional consensus around what a new teacher should be able to do. With no common language for describing and analyzing teaching, we have a weak basis for a system of training and assessing teaching practice. This is the case across the entire enterprise of teacher training and development, from traditional, higher education-based programs to those run by school districts and non-profit organizations." See http://www.teachingworks.org/training/seminar-series.

specific theoretical or practical knowledge, three institutionalized alternatives present themselves to network executives as potential strategies for organizing the development and scale up of school-wide improvement models: "shell enterprises" that seek to replicate distinguishing organizational characteristics (e.g., roles, structures, culture) absent earnest efforts to replicate capabilities; "diffusion enterprises" that seek see to codify established practices to be enacted with fidelity in schools; and "incubation enterprises" that provide principles and parameters to structure and constrain school-based design and problem solving.

Each is potentially viable under particular conditions. For example, shell enterprises can be viable when isomorphism (rather than effectiveness) is sufficient to secure legitimacy and resources. Diffusion enterprises can be viable when the hub succeeds in appropriating sufficient practical knowledge to ensure effectiveness, and when individual schools present neither exceptional circumstances nor a desire to exercise agency and discretion. And incubation enterprises can be viable when the hub succeeds in identifying schools with existing capabilities for design and continuous improvement, and when there is no need to link school-level effectiveness to a consistent "treatment."

However, these do not appear to be the conditions under which most school improvement networks currently operate: pressed beyond establishing innovative shells to demonstrating replicable effectiveness; absent established professional knowledge and practices to diffuse; and (to the extent that they heed policy pressure and incentives) working with schools lacking the capabilities for design and continuous improvement that are needed to support incubation. Moreover, as a general matter, the problems of each of these strategies are well-established, and long associated with enduring problems of US education reform that risk undermining summative impact evaluations.

For example, shell strategies have been associated with loose coupling and non-implementation (Meyer & Rowan, 1978; Meyer, Scott, & Deal, 1983). Diffusion strategies have been associated with technocratic and/or bureaucratic dispositions, rote compliance, and unresponsiveness to local circumstances (Berman & McLaughlin, 1975, 1978; Peurach, 2011). Incubation strategies have been associated with individual and organizational autonomy, program cooptation, and regression to past practice (Firestone & Corbett, 1988; Leithwood & Menzies, 1998; Muncey & McQuillan, 1996). And, as external initiatives, all of these strategies have been bound up with problems of confusion, politics, rejection, and/or abandonment.

Consequences and Considerations for Evaluating School Improvement Networks

If the objective of conventional impact evaluations is "to understand what works, what doesn't, and why," then the preceding analysis suggests that those making decisions to support, fund, and/or enlist in school improvement networks based on evidence of replicable effectiveness are working under conditions of tremendous uncertainty (not to mention those seeking to operate these networks). One problem is that the evolution, variation, and confounding of "subjects,", "treatments,", and "administrators" (compounded by fragmented, turbulent, and weak environments) greatly complicates efforts to discern the effects of school-wide improvement models on student achievement (never mind discerning the underlying causal dynamics). A second -- and related -- problem is that the highly institutionalized logic that underlies the conventional evaluation regime (the diffusion-centered, RDDU logic) appears to be at odds with the ways in which school improvement networks actually emerge and mature.

Indeed, recognition of the above-described challenges and realities has contributed to efforts to fundamentally reframe understandings of the processes by which these networks emerge and mature. For example, rather than some fixed, objective "treatment," researchers

have reconceptualized school-wide improvement programs as subjective realities created through processes of co-construction and sensemaking among schools, districts, program providers, and other vested organizations (Datnow, Hubbard, & Mehan, 2002; Datnow & Park, 2009). Further, researchers describe such work as requiring both exploiting available knowledge and exploring new directions (Hatch, 2000); as requiring that schools take ownership in order to effect both deep and broad change in core practices, understandings, and values (Coburn, 2003); and as fraught with challenges and puzzles (Cohen et. al, in press; Hatch & White, 2002; McDonald, Klein, & Riordan, 2009). Finally, rather than emerging through RDDU-like processes, researchers have re-conceptualized the process by which school improvement networks develop and mature as a set of interdependent functions enacted concurrently and iteratively by hubs and schools over time. Examples of these processes include obtaining funding; designing and improving programs; recruiting schools; supporting implementation; evaluating effects; and building capacity in the hub (Farrell, Nayfack, Smith, Wohlstetter, & Wong, 2009; Glennan, Bodilly, Galegher, & Kerr, 2004; Peurach, 2011).

Extending the preceding, Peurach and Glazer (2012) argue that such work is best understood when examined *not* through the lens of a diffusion-oriented logic but, instead, through the lens of an evolutionary logic in which hubs and schools collaborate over time to produce, retain, use, and improve, a formal knowledge base supporting replicable effectiveness. The evolutionary logic, in turn, bears close resemblance to other methods of design-based implementation research in education (Penuel, Fishman, Cheng, & Sabelli, 2011). Further, in a longitudinal, quasi-experimental study of three leading comprehensive school reform strategies, two hub organizations using an evolutionary strategy (Success for All and America's Choice) demonstrated positive, significant, and replicable effects in improving leadership, instruction,

and student achievement, with those outcomes attributed to extensive, formal supports for instructional practice and for teachers' practice-based learning (Camburn, Rowan, & Taylor, 2003; Rowan, Correnti, Miller, & Camburn, 2009a; 2009b). A three-year, randomized field trial of Success for All also showed positive, statistically significant, and replicable program effects on student achievement (Borman, Slavin, Cheung, Chamberlain, Madden, & Chambers, 2007).

To be clear, none of the preceding is to argue away the need for rigorous evaluation of replicable effectiveness. Given the billions of dollars in play, and given the high stakes for children (indeed, for society as a whole), there is a clear imperative for rigorous impact evaluations, especially those that go beyond identifying main effects to examine causal dynamics. After all, to launch a school improvement network (or, for that matter, any educational reform) is to experiment on and with children. Moreover, some of the challenges described above can be managed with sophisticated research designs, complex statistical procedures, and very large samples sizes -- though at the expense of increasing the demands on the scarce resources and capabilities of hub organizations, networks, evaluators, and environments. Regression discontinuity designs (a widely prescribed antidote to the above-described challenges) are a case in point (Schochet, 2008).

Instead, the preceding analysis is intended to support three points. The first is that answering the questions "does the program work?" and "can success be replicated?" is a long-term, expensive, and uncertain undertaking. The second is that that this uncertainty leaves networks vulnerable to practical and methodological issues that complicate meeting standards for replicable effectiveness. The third is that such challenges and problems strongly suggest the need for complementary, formative evaluations anchored deeply in what researchers are learning about the ways that networks emerge, evolve, and mature over time.

Developmental Evaluation: Logic, Criteria, and Considerations

If the goal, ultimately, is to conduct summative impact evaluations that establish causality, then the one aim of formative developmental evaluation should be to assess the emergence and maturation of "that which causes": specifically, knowledge supporting replicable effectiveness. Indeed, a central tenet of contemporary educational reform is that prospects for increasing student achievement do not lie primarily in improving roles, structures, resources, and culture in schools but, instead, in improving the practices and understandings of teachers and school leaders as they construct, enact, and manage instructional and non-instructional services for students. And, as argued, a central problem of contemporary educational reform is the shortage of precisely such knowledge, in schools and in their environments.

As such, we continue by reviewing the evolutionary logic of replication detailed by Peurach and Glazer (2012) in order to propose five criteria (and associated considerations for interpretation) to structure the developmental evaluation of school improvement networks.[8] As described above, the logic provides a way of thinking and reasoning about school improvement networks as producing, retaining, using, and improving practical knowledge through collaborative, experiential learning among hubs and schools. The logic was originally drawn from leading theory and research on franchise-like organizational replication in the commercial sector, proposed as an ideal type for interpreting and comparing school improvement networks, and used to construct an interpretation of one leading school improvement network (Success for All) as a knowledge-producing enterprise.[9]

---

[8] We originally termed this a "knowledge-based logic" of replication. Through subsequent exchanges with Sidney Winter, we came to recognize that one of our primary critical foils (the RDDU sequence) is, itself, a knowledge-based logic. Hence, our shift to referring to this as an "evolutionary logic", out of recognition of the logic's deep roots in evolutionary economics.

[9] As noted in our earlier synthesis (Peurach and Glazer, 2012), the evolutionary logic is drawn from theory and research by Sidney Winter, Gabriel Szulanski, and colleagues, much of it rooted in the Wharton School at the University of Pennsylvania, and much of it focused on the replication of knowledge within and between

Review: The Evolutionary Logic of Replication

As with school improvement networks, the evolutionary logic begins with a central, hub organization replicating a common organizational design across large numbers of outlets. The organizational design is assumed to be sufficiently broad in scope as to transform the core capabilities (and even the identity) of outlets, with the goal of replicating the effectiveness of production activities and/or service delivery (Winter & Szulanski, 2001). The chief mechanism of replication is formalized, codified knowledge intended to enable (rather than coerce) production and/or service delivery in outlets (Adler and Borys, 1996). Using Success for All as an education-specific example, the "hub" would be the independent, non-profit Success for All Foundation (SFAF). The "outlets" would be the individual schools with which SFAF works. And the organizational design would be the Success for All program.

Such a strategy has advantages in terms of speed, efficiency, and effectiveness over outlet-by-outlet invention under two conditions. The first is when conditions limit the straightforward appropriation or acquisition of essential knowledge (e.g., weak professional knowledge, education, and human resources in environments and in outlets). The second is when conditions limit the social retention and reproduction of essential knowledge through apprenticeship, mentoring, and communities of practice (e.g., long distances between hubs and outlets; high ratios of outlets to templates; and personnel transiency).[10] Straightforwardly, if

---

organizations: for example, Baden-Fuller and Winter (2005); Szulanski & Winter (2002); Szulanski, Winter, Cappetta, & Van den Bulte (2002); Winter (2003, 2010, 2012); Winter & Szulanski (2001, 2002); and Zollo & Winter (2002). The basis of this work lies in the work of Nelson & Winter (1982) on evolutionary economics, with specific focus on developing, adapting, and replicating routines. The perspective has contemporary ties to research in: organizational learning (March, 1991/1996); innovation development (Van de Ven, Polley, Garud, & Venkataraman, 1999); organizational routines (Feldman & Pentland, 2003); dynamic capabilities, the resource-based view of the firm, and the evolutionary view of the firm (Arrow, 1962, 1974; Brown & Duguid, 1998; Eisenhardt & Martin, 2000; Grant, 1996; Wernerfelt, 1995); alternative conceptions of centralized control (Adler & Borys, 1996); franchised organizational forms (Bradach, 1998); and non-profit replication (Bradach, 2003).

[10] See Baden-Fuller and Winter (2005) on conditions supporting replication via "principles" (which we refer to as an incubation strategy) and replication via "templates" (which we refer to as an evolutionary strategy).

essential knowledge is either weak or non-existent, and if it is difficult to retain and share knowledge person-to-person and organization-to-organization, then it becomes incumbent upon the hub both to produce and retain essential knowledge and to devise other means of recreating it in outlets.

Premises: Practice-Focused, Learning-Driven Networks.

The evolutionary logic begins with two core premises. The first premise is that, in replicating complex organizational models, the overarching consideration is *not* the replication of roles, structures, or culture, simply because it is possible to replicate broad organizational forms without replicating organizational effectiveness (Winter & Szulanski, 2001).[11] Instead, the overarching consideration is the replication of capabilities: that is, the replication of practices and understandings that support working differently, more effectively, and in more coordinated ways to effect intended outcomes.

The second premise is that capabilities cannot be reliably replicated through the rapid, unilateral transfer, communication, or dissemination of knowledge and information from hubs to outlets, owing to uncertainties (and potential shortcomings and flaws) in available knowledge, inaccuracies and uncertainties in communication, and the complexities of human agents learning to enact and understand their work in new ways. Instead, the evolutionary logic holds that the replication of organizational capabilities requires the *creation* and *recreation* of coordinated, interdependent practices and understandings through collaborative, experiential, long-term learning among hubs and outlets.

Foundations: Essential Knowledge Base and Core Learning Processes

[11]Akin to our notion of a "shell" enterprise, replicating organizational apparatus without replicating capabilities is described in terms of "faux replication" (Winter and Szulanski, 2001; 2002). Consistent with notions of isomorphism as developed by DiMaggio and Powell (1991), faux replication can be adaptive when hubs seek to benefit from franchise fees in the short term without incurring the high costs of developing capabilities, and/or when outlets seek to exploit the reputational assets of the network.

Given the preceding, the primary focus of the evolutionary logic is the production and use of an essential knowledge base that supports the broad scope replication of capabilities. This knowledge base consists of three categories: knowledge of *what*, *how*, and *where* to replicate (Winter and Szulanski, 2001). Knowledge of *what* to replicate focuses on the essential practices and understandings to be recreated in each outlet. Knowledge of *where* to replicate focuses on practices and understandings within the hub for identifying, vetting, and selecting outlets and environments that favor successful replication. Knowledge of *how* to replicate focuses on practices and understandings within the hub for recreating essential practices and understandings in outlets (e.g., strategies for training and coaching).

This essential knowledge base is generated, reproduced, used, and refined through multiple iterations of two interdependent learning processes co-enacted by hubs and outlets: exploitation and exploration (Winter & Szulanski, 2001; see, also, Bradach, 1998, and March, 1991/1996). Exploitation is the process of leveraging available knowledge in new contexts and learning from experience.[12] Exploration is the process of identifying new possibilities for what, where, and how to replicate through search, experimentation, discovery, and invention.

Emergence: A Template

To establish proof of concept, development of the essential knowledge base begins with the construction of a "template": a working example (or examples) of the production or service capabilities to be replicated, often constructed in carefully selected sites with carefully selected people (Baden-Fuller & Winter, 2005; Winter, 2010; Winter & Szulanski, 2001). The template functions as a context for initial, exploratory learning in which hub and template staff engage in

[12] The connotation of "exploitation" is entirely positive (and not negative), as in "making full use of" (in contrast to "benefitting unfairly from").

joint search, experimentation, discovery, and invention to devise means of realizing intended ends.

With successful exploration, the template becomes a repository of tacit knowledge from which the hub can begin developing understandings of what capabilities are to be recreated in outlets, where those capabilities might be recreated, and how to recreate them. It also functions as a resource for developing a formal design for practice: a description of essential roles; operating principles detailing responsibilities associated with each role; and first principles that structure and coordinate outlet-wide activity.

Essential Resource: Formalized Knowledge

With proof of concept, a central role of the hub is to formalize the essential knowledge base: that is, to codify knowledge of what, where, and how to replicate in manuals, training materials, digital media, tools, and other artifacts (Winter and Szulanski, 2001; 2002).[13]

Formalized knowledge takes two forms. The first form is codified routines: coordinated patterns of activity, both in outlets (e.g., routines supporting essential practices) and in the hub (e.g., routines supporting the selection and creation of outlets). These include "closed" routines: procedures that provide step-by-step directions for what, exactly, to do in particular situations. They include "open" routines: frameworks used to devise courses of action under conditions of uncertainty. They include assessment routines used to generate information with which to evaluation performance and outcomes. And they include "learning" routines that detail cycles of diagnosis, planning, implementation, and reflection. Routines are considered the primary

[13] The work of Winter, Szulanski, and colleagues generally places more emphasis on routines than on guidance. However, the importance of professional and background knowledge becomes salient in Baden-Fuller & Winter (2005) as a complement to routines. Moreover, we developed this notion in our earlier synthesis under the topics of "supplemental guidance" and "information resources" as complements to routines. Note that subsequent consideration has us reconceptualizing our prior notion of "information resources" as "assessment routines".

mechanisms for supporting levels of coordinated activity that would otherwise be difficult and costly to achieve (Nelson and Winter, 1982).

The second form is codified guidance to support responsiveness to local circumstances and exigencies, the management of inevitable breakdowns and limitations in routines, and the intelligent (rather than rote) selection and enactment of routines. Beyond a formal design for practice, such guidance can include professional and background knowledge essential to the enactment of specific roles and responsibilities; goals and standards for performance; and evaluation rubrics and decisions trees that support analysis and decision making.

Endemic Complication: Partial and Problematic Knowledge

Within the evolutionary logic, an endemic complication is that the hub often faces pressure to begin scaling up before having a completely worked out template or a highly developed formal knowledge base (Winter & Szulanski, 2001). Within the template, activities may combine to effect intended outcomes in non-obvious ways; relevant knowledge may remain tacit; understandings of cause-and-effect relationships can be flawed; and apparently-important activities may be completely unrelated to outcomes. Further, the effectiveness of templates is likely to depend on specific individuals, relationships, and environments in ways not fully understood at the outset.

Consequently, consistent with established understandings of satisficing, hubs and outlets typically commence replication with potentially-rich (but partial-and-problematic) knowledge of key practices and understandings to be replicated in outlets, and with only emergent knowledge about where and how to replicate them. Consider an alternative (and unlikely) case: the possibility that, working from one or a small number of templates, the hub would be able to quickly discern and formalize perfect knowledge of what, where, and how to replicate.

Essential Method: Developmentally-Sequenced Replication

The evolutionary logic continues with the hub recruiting or developing outlets and proceeding to large-scale replication, with the goal of recreating conventional capabilities for achieving common performance levels across outlets. The method for doing so is a developmentally-sequenced replication process that depends on a synergy between two approaches to replication often viewed as logical opposites: fidelity of implementation and adaptive, locally-responsive use (Szulanski, Winter, Cappetta, & Van den Bulte, 2002; Winter, 2010; Winter & Szulanski, 2001).[14] Consistent with exploitation as a core learning process, the former focuses on recreating established practices and understandings in new outlets in ways that mirror conventional understandings of diffusion. Consistent with exploration as a core learning process, the latter focuses on extending and refining those practices and understandings in ways that mirror conventional understandings of incubation.

The developmental sequence begins with fidelity of implementation: enacting formalized routines as specified, with the goal of establishing conventional, coordinated, base-level capabilities and performance levels within and between outlets. Despite shortcomings and problems in the essential knowledge base, and despite the deferred benefits of addressing outlet-specific exigencies, fidelity of implementation provides multiple advantages: for example, mitigating against weak initial capabilities in outlets; taking advantage of lessons learned and problems solved; learning by doing (e.g., to enact new practices, to understand underlying principles, and to understand the interdependence and coordination of activities); forestalling

[14] As noted in our earlier synthesis, Szulanski, Winter, Cappetta, & Van den Bulte (2002) actually cast this as a four-phase process. *Initiation* involves recognizing opportunities to replicate and deciding to act on them. *Initial implementation* is a process of "learning before doing," either by planning or by experimenting before actually putting knowledge to use. *Ramp up* to satisfactory performance is a process of learning by doing and of resolving unexpected outcomes. Finally, *integration* involves maintaining and improving the outcome of the transfer after satisfactory results are initially obtained. Thus, initiation, initial implementation, and ramp focus on exploitation, and have, as a core focus, fidelity of implementation. Integration begins to introduce experimentation and has, as a core focus, local adaptation.

early problems (e.g., regression to past practice; the introduction of novel, site-specific operational problems); and establishing conventions that support collaborative learning and problem solving (e.g., common language, shared experiences, and joint work).

Once base-level practices and understandings are established, the developmental sequence proceeds to adaptive use. With that, outlets assume ownership and assert agency in enacting the model in order to compensate for shortcomings, address problems, and respond to local needs and opportunities. Adaptive use can include adjusting hub-formalized routines and guidance to better address local circumstances; inventing new routines and guidance that address critical work not yet formalized by the hub; and/or abandoning routines and guidance that appear either inconsequential or detrimental.[15] Capabilities for adaptive use are not assumed. Rather, the hub supports such activity using open routines that support local decision making; assessment routines for evaluating performance and outcomes; "learning routines" that guide analysis, evaluation, and reflection; and guidance that provides knowledge, goals, standards, and information that both support and constrain local analysis, invention, and problem solving.

The enactment of this developmental sequence also creates opportunities for the hub to engage in its own learning in order to leverage, extend, and refine knowledge of where and how to replicate. Regarding where to replicate, this involves enacting and adapting routines and guidance for identifying outlets and environments that meet pre-conditions for initial, faithful implementation and, with experience, are prepared to advance to adaptive use. Regarding how

[15] As an education-specific example, this might include incorporating district-required literacy modules and assessments into a comprehensive, externally-developed curriculum; devising remedial self-study modules for students struggling with particular content in that curriculum; and/or selectively eliminating a subset of instructional tasks that addresses particular content in ways that appear at odds with state accountability assessments.

to replicate, this involves enacting and adapting and routines and guidance for working with outlets to develop capabilities both for base-level operations and for adaptive use.

The Outcome: Knowledge Evolution

This developmental sequence fuels a knowledge evolution cycle through which the hub and outlets collaborate to continuously expand and refine the essential knowledge base (Zollo & Winter, 2002). The cycle begins with fidelity of implementation within and between outlets to establish conventional, base-level capabilities and performance levels. As they advance to adaptive use, outlets introduce variation into the network regarding practices and understandings that support effective operations. As the coordinative center, the hub monitors the network for instances and patterns of variation; selects and evaluates potential improvements; squares those with existing or new knowledge, resources, and requirements in broader environments; and retains improvements both by incorporating them into an evolving template and by formalizing them as routines and guidance. New practices and understandings are then fed back into the installed base of outlets as incremental, "small-scope" improvements, and they are incorporated into a broader-yet knowledge base to support the creation of new outlets.

The cycle then begins again, with initial recreation of practices and understandings via faithful implementation, followed by adaptation, variation, selection, and retention. Successive iterations result in an increasing (and increasingly refined) formal knowledge base detailing where, what, and how to replicate.

Essential Mechanism: Dynamic Capabilities

Such iterative knowledge evolution is highly dependent on dynamic capabilities through which hubs and outlets systematically generate and modify practices and understanding in

pursuit of improved effectiveness, continued legitimacy, and sustainability (Dosi, Nelson, & Winter, 2001; Winter, 2003; Winter & Szulanski, 2001; Zollo & Winter, 2002). In outlets, dynamic capabilities are anchored in the sort of adaptive use described above. In hubs, dynamic capabilities are anchored in infrastructure and capabilities for rapidly pooling and analyzing information and knowledge from throughout the network; for evaluating the relationship between practices and understandings (on the one hand) and intended outcomes (on the other); for experimentation, rapid prototyping; and in goals, standards, and capabilities; and for disseminating program improvements through the installed base of outlets.

Extensive iterations will not yield omniscience. The essential knowledge base will always be partial and problematic, and key knowledge will always remain undiscovered and/or tacit. As such, knowledge evolution - featuring cycles of exploitation and exploration - functions as the essential capability of network-based organizational replication initiatives, enacted jointly by hubs and outlets over the life of the enterprise to support base-level operations, adaptive use, continuous improvement, and long-term viability.

Criteria for Developmental Evaluation

Thus, from the perspective of the evolutionary logic, the question driving the developmental evaluation of school improvement networks would not be, "Does the program work?" Rather, the driving question would be, "Is the enterprise *working* in ways likely to yield a formal knowledge base supporting the large-scale replication of capabilities?"

Continuing to draw on the knowledge-based logic, we adapt and extend criteria first proposed by Peurach and Glazer (2012) and Peurach, Glazer, and Lenhoff (2012) as having potential to provide evidence that a school improvement network is (or is not) developing and functioning in ways consistent with the evolutionary logic. While not exhaustive, these five

criteria have potential to structure the collection of a parsimonious-yet-powerful body of evidence for use by funders, hubs, schools, and other vested parties in considering progress toward developing the logical antecedents to successful impact evaluation: formal knowledge of where, what, and how to replicate.

The first criterion is an initial, screening question intended to determine the appropriateness of evaluating a given school improvement network as an evolutionary enterprise. Assuming conditions warrant evaluation as an evolutionary enterprise, the following four criteria examine features of the network with potential to support the production, retention, use, and refinement of a formal knowledge base.

1. Do conditions warrant developmental evaluation as an evolutionary enterprise? Such conditions include limitations on the social retention and reproduction of knowledge: for example, long distances between the hub and schools; high ratios of outlets to templates; high ratios of school staff to hub training staff; and personnel transiency. Such conditions also include limits on straightforwardly appropriating or acquiring essential knowledge to support goals for school-wide improvement: for example, as evidenced by the absence of reviews and meta-analyses of research; of established resources and methods for enacting essential practices (e.g., favorable reviews in the What Works Clearinghouse or Best Evidence Encyclopedia); of agencies and organizations chartered with evaluating and synthesizing essential knowledge (e.g., the National Reading Panel); and of organizations and agencies that provide pre-service and in-service professional education to support essential practices and understandings.

2. Does the enterprise have a replication infrastructure? Such an infrastructure is evidenced by a formalized design for practice (i.e., descriptions of essential roles, along with principles detailing responsibilities and the coordination among them); an operating template

that functions as proof of concept; and an explicit strategy for replication that combines exploitation *and* exploration in ways that support the evolution of a formal knowledge base.

3. Does the enterprise feature formal, codified resources for recreating base-level capabilities in outlets? These resources would be evidenced by formal routines and guidance for recruiting, selecting, and enlisting outlets in which conditions exist (or can be created) to support base-level operations; by formal routines and guidance for use by outlet staff to establish consistent, base-level practices and understandings; and by formal routines and guidance for use by trainers and coaches to support outlets in establishing base-level practices and understandings.

4. Does the enterprise feature formal, codified resources for recreating capabilities for adaptive, locally-responsive use? These resources would be evidenced by formal routines and guidance for use by hub staff in identifying outlets that have capabilities for base-level operations (and, thus, are prepared to progress to adaptive use); by formal routines and guidance for use by outlet staff to support design, evaluation, problem solving, decision making, and other discretionary activity; and by formal routines and guidance for use by trainers and coaches to support such activity in outlets.

5. Does the hub organization have the infrastructure and capabilities to support evolutionary learning? Such infrastructure and capabilities are evidenced by the above-described supports for adaptive, locally-responsive use as a source of within-network variation in practices and understandings; a communication infrastructure supporting the reciprocal exchange of knowledge and information among hubs and schools; opportunities, resources, and capabilities in the hub for analysis and problem solving (including formal goals and standards for analyzing performance and outcomes in outlets); opportunities, resources, and capabilities in

the hub for rapidly prototyping, evaluating, and formalizing new resources; and mechanisms for disseminating new resources through the installed base of schools (e.g., the above-described capabilities for supporting base-level operations).

Considerations for Analysis

In considering their use in analysis, one conjecture is that more strengths across more of the proposed criteria would increase the potential for a network to function in ways consistent with the evolutionary logic. The corollary is that more weaknesses across more criteria would increase the risk of the "Matthew effect" or "digital divide" long common in education reform, with existing absorptive capacity and dynamic capabilities predicting implementation and outcomes. That is, schools that enter a network with prior capabilities (both for practice and for *learning from* practice) would have potential to leverage hub-provided resources to improve. Schools that enter a network lacking such capabilities would be susceptible to enduring problems of externally-sponsored education reform, all of which compromise the treatment in ways likely to undermine summative impact evaluations: non-implementation, owing either to confusion, rejection, or abandonment; rote compliance, absent attention to effectiveness or to local exigencies; unconstrained adaptation, resulting in cooptation and/or regression to past practice; or some combination, within and between staff members and program components.

Three additional considerations should further mediate the use of these criteria. The first is that, given that understandings of the evolutionary logic are nascent as compared to the institutionalized alternatives, it is unlikely that a given school improvement will have intentionally elected to pursue an evolutionary strategy. Even so, it is possible that the network is poised to "evolve to evolve," with the hub and schools learning of the need and possibility to combine shell, diffusion, incubation, and (possibly) other, yet-to-be devised strategies in novel

ways to support network-wide learning and improvement.[16] In fact, the notion of developmental evaluation is premised on precisely that possibility.

The second is that developing in ways consistent with the evolutionary logic does not imply smooth sailing. In fact, development as an evolutionary enterprise actually has potential to introduce steep challenges into the network: for example, designs for practice that intervene on historically private and autonomous work; routines and guidance for base level operations that could easily be interpreted either as bureaucratic interventions or as technocratic quick fixes; routines and guidance for adaptive use that could easily be interpreted as license to do one's own thing; and constantly-improving program resources that resemble the usual environmental churn.

Finally, it is important to recognize that the proposed criteria operate at a high level in order to examine what we view as the foundational elements of an evolutionary enterprise. Complementary analyses would be needed to examine the content of routines and guidance; the actual use of program resources in schools; and the work of hubs in leveraging school-level adaptations as resources network-wide improvement. Thus, the proposed criteria should be understood as a first step toward developmental evaluation, and not the whole story.

A Developmental Evaluation of the New Tech Network

To investigate our proposed criteria, we apply them to a developmental evaluation of the New Tech Network, a school improvement network in which a hub organization is working to replicate a school-wide design for project-based learning in more than 100 high schools across the country. In 2012, the network was awarded a $3 million i3 development grant to support

[16] For example, in two cases documented as operating as evolutionary enterprises (Success for All and America's Choice), the evolutionary approach was less an intentional, explicit strategy and more a pragmatic, tacit strategy, with hubs that were aggressively pursuing either a diffusion or incubation strategy learning over time to combine the two in support of both conventional, base-level operations and adaptive, locally-responsive use (Cohen et. al, in press; Peurach, 2011; Peurach and Glazer, 2012).

two STEM-focused high schools in South Carolina. Below, we provide additional background on the New Tech Network, after which we report our research procedures, findings, and possible topics for formative conversations among stakeholders. In our view, this study provides evidence of the potential power of our proposed criteria for providing formative feedback to funders, hubs, and schools regarding strengths and weaknesses in their network as they progress together toward summative impact evaluation.

The New Tech Network

Headquartered in Napa, CA, the New Tech Network is a non-profit school improvement network that operates as a subsidiary of the KnowledgeWorks Foundation of Cincinnati, Ohio. For 2012/2013, the network will include 125 schools in 19 states (118 high schools and seven middle schools): for sake of comparison, more high schools than supported by seven state education agencies, and roughly as many high schools as in the states of Maine and Nevada (New Tech Network, 2012a).[17] The network includes both established and newly-created schools (both freestanding and "schools-within-a-school").

For 2012/2013, fees for the initial, 4.5 year contract are between $450,000 and $500,000, with continuation fees estimated at $20,000 per year. Among other materials and services, these fees cover access to *Echo*, the New Tech Network's online learning management system. They also cover coaching and conference costs that include five days of initial training in the summer preceding Year 1 implementation; a minimum of seven days of site-based support from a New Tech coach; approximately two weeks per school of facilitated collaboration among groups of geographically-proximal schools; two two-day leadership summits; and an annual three-day conference (New Tech Network, 2011; New Tech Network, 2012b).

[17] For the number of high schools per state, see Williams, Blank, Toye, and Petermann (2007).

The New Tech Network is not the creation of a seasoned hub organization with extensive experience supporting large-scale, school-wide improvement. Rather, one hub staff member described the network as a "homegrown" enterprise, with the hub, schools, and program co-emerging over a sixteen year period, in interaction with the rise of high school reform on the national agenda and in ways consistent with the evolutionary logic.

The overarching goal of the New Tech Network is "to enable students to gain the knowledge and skills they need to succeed in life, college and the careers of tomorrow" (New Tech Network, 2012c). This goal was initially articulated in terms of "21st century skills": e.g., critical thinking, oral communication, collaboration, and creativity. Subsequently, it has been articulated in terms of "deeper learning" and "college and career readiness".

Toward these ends, New Tech features a school-wide improvement program with three core elements (New Tech Network, 2012d). The first is a common design for interdisciplinary, project-based learning intended to transform schools' core instructional capabilities in all academic content areas. The Buck Institute for Education (which New Tech identifies as a chief resource for program development) describes project-based learning as "an extended process of inquiry in response to a complex question, problem, or challenge. While allowing for some degree of student 'voice and choice,' rigorous projects are carefully planned, managed, and assessed to help students learn key academic content, practice 21st Century Skills (such as collaboration, communication & critical thinking), and create high-quality, authentic products & presentations" (Buck Institute for Education, 2012). The second is extensive use of information technology, including one-on-one student computing. The third is a focus on establishing a culture of trust, respect, responsibility, and accountability. These three core elements are

complemented by a focus on establishing external partners to support implementation and effectiveness, including local businesses, colleges, universities, and government agencies.

The first New Tech school, Napa New Technology High School, was established in 1996 in the Napa Valley Unified School District, the product of a four-year effort by education, business, and community leaders to re-imagine high school education (Borja, 2002). In 2003, supporters secured a $6 million replication from the Gates Foundation to establish the New Tech Foundation, with the goal of developing 14 new schools in a three year period. Continued philanthropic support, the acquisition by KnowledgeWorks, and movement to a fee-for-service financial strategy fueled continued growth: expansion to 40 schools by 2009/2010, followed by the addition of 85 new schools between 2010/2011 and 2012/2013 (a three year growth rate of 313%).[18]

This growth was described by one NTN staff member as "more serendipitous than planful," and driven by available funding, schools' interest, and internal ambitions. It also brought increasing diversity to the initial capabilities and environmental contexts of New Tech schools: at the one extreme, the initial, self-created high school in Napa, CA; and at the other extreme, New Tech's i3-funded high schools in South Carolina, in what it describes as "two of the nation's persistently lowest-achieving, lowest income, most economically under-resourced rural communities" (Furman Institute, 2012). As of SY 2010/2011, 37% of schools were in urban districts, 25% in suburban districts, and 38% in rural districts. Further, 50% of students were female, 57% were of color, 50% were eligible for free or reduced priced lunch, and 5% were English language learners (New Tech Network, 2012e).

[18] In addition to contributions from the Gates Foundation and KnowledgeWorks, and in addition to fees from schools, the New Tech Network reports that growth in the enterprise benefitted from financial support and technical assistance from (among others) the Carnegie Corporation of New York, the Hewlett Foundation, the Steelcase corporation, the Toshiba corporation, the Buck Institute for Education, and several state-level reform enterprises. See http://www.newtechnetwork.org/newtech_partners.

In the ten years since its founding, the New Tech hub has expanded to an estimated 45 total staff members, 15 in the central office in Napa and 30 who serve as field-based development and training staff.[19] The hub is currently organized into six primary units: executive leadership; program leadership; school design and implementation; new school development and planning; technology development and support; and community, innovation, and research.

The New Tech Network is a strong candidate for developmental evaluation. To date, we could not identify any rigorous internal or external evaluations showing statistically significant, replicable program effects on student outcomes as compare to non-New Tech schools.[20] Moreover, the combination of continued growth, increasing prominence, and continued public and private investment is likely to soon draw pressure to demonstrate replicable effectiveness on summative impact evaluations.

While they view their work to date as a success, hub staff members also recognize the prospects of summative impact evaluation. In a 2012 interview, one executive explained that "we need to be able to demonstrate that the work we do can be replicated -- that we can maintain quality and reproduce the same impact, the same results, in a myriad of communities, types of schools, and types of students." Also in a 2012 interview, another staff member explained that the issue thus becomes one of replicating capabilities across schools: "It's really

[19] Staffing estimates and organizational structure are taken New Tech interviews conducted during the winter of 2012 and from New Tech Network (2012f).

[20] While searches of conventional databases revealed position papers and other commentary on the New Tech Network, we did not identify any peer-reviewed studies of program implementation or effectiveness as compared to non-New Tech Schools (and, thus, no meta-analyses or best evidence syntheses of such studies). The primary source of evidence is the network's own web site, which provides links to a collection of internal and external documents and studies of implementation and effectiveness. See http://www.newtechnetwork.org/newtech_results. While most of the information alludes to high level student outcomes (though often absent comparisons to demographically comparable non-program students), some actually suggests ACT and SAT scores that fall below national averages. Quantitative analyses are complemented by a small number of school-specific case studies, most of which (again) are internal reports or self-reports alluding to the potential and the promise of the program.

hard for people to take all this great technical expertise and know how to place it into a school and then use it as a tool… We have coaches and people in our organization, the 'why' and the purpose is in their bones. It is now tacit for them. It is a part of who they are. And, so, how do we stop and make sure that it becomes a part of who these new school leaders are so they can build that in teachers?"

Research Procedures

We conducted our developmental evaluation in the context of a broader study examining efforts within the New Tech Network to improve instructional practice concurrent with building the educational infrastructure needed to do just that: a challenge endemic to instructionally-focused school improvement networks as a strategy for large-scale education reform.[21]

Study Design.

Our study design derives from experience conducting ethnographic case studies of leading comprehensive school reform programs (Cohen et. al, in press; Peurach, 2011). Specifically, we designed our analysis as an exploratory case study using a longitudinal, embedded case study design (Scholz & Tietje, 2002; Stebbins, 2001; Yin, 2009). The New Tech Network functions as the case. Within the network, we examined three distinct sub-units and the relationships among them: the New Tech hub organization, the New Tech school-wide improvement model, and three New Tech schools that began implementation in SY 2010/2011 (all within the same state, though each with a unique student, staff, and geographic context). Consistent with understandings of a "community infrastructure" supporting school improvement networks (Glazer and Peurach, 2012), we examined this case as situated in a broader environmental context consisting of four key components: policy, regulatory, and other institutional supports; resource endowments; market functions; and proprietary activity.

[21] Results from the broader study are forthcoming.

Data Collection

Data collection spanned two years, May, 2010 to May, 2012. Consistent with methods of organizational ethnography (Brewer, 2004; Fine, Morrill, & Surianarain, 2008; Lee, 1999), it included the collection of documents and artifacts, participant-observation, and interviews. Besides collecting training materials, instructional materials, and research reports, we secured access to (and regularly reviewed) *Echo*, the New Tech Network's online learning management system and repository of thousands school-created projects, hub-created projects and guidance, and other supporting materials.

Further, as participant-observers, we participated in six day-long site visits in each of the three schools; two statewide professional development sessions; four national conferences; and nine formal and informal meetings between New Tech leaders, district coordinators, and New Tech staff members. We also conducted two sessions at the New Tech Networks annual conference in the summer of 2011 that were focused on fostering conversation among hub and school staff about possible synergies between fidelity and adaptation, and we collaborated with a regional education service agency to co-facilitate a standing "professional learning community" composed of the directors of New Tech schools in one state.

Finally, we conducted 20 semi-structured interviews with 17 participants involved in the implementation of the New Tech programs in the three schools participating in our study: two superintendents, two school directors, ten teachers, two regional district coordinators, and one New Tech school development coach); and document and artifact collection from New Tech and school personnel. In addition, we also conducted eight semi-structured interviews with staff members in the New Tech hub, including executives, lead developers, and lead trainers (including staff members who have been with the network since its inception). We

complemented our interviews with ongoing, informal conversations with staff members from the New Tech hub, our participating schools, and one regional educational service agency both to learn more and to provide feedback.

Analysis

We used iterative memo writing as our primary analytical method (Miles and Huberman, 1994), concurrent with (and in interaction with) our data collection, and with explicit attention to leveraging principles of positive organizational scholarship in maintaining a empathetic-yet-critical stance in seeking to identify and report strengths and vulnerabilities within the network (Cameron and Spreitzer, 2011; Dutton, Quinn, and Cameron, 2003). For our broader study, this involved categorizing and reporting evidence about schools, the program, the hub organization, and broader environments. For the developmental evaluation, this initially involved categorizing and reporting evidence using questions first proposed by Peurach and Glazer (2012) and subsequently refined by Peurach, Glazer, and Lenhoff (2012).

For the developmental evaluation, given our analytic focus on formal resources, the general pattern was to analyze documents and artifacts; observe their use; and discuss them, their origins, their use, and their evolution with New Tech and school staff. Multiple iterations of analysis and data collection drove clarifications in our exposition of the evolutionary logic (detailed above). Further, given that the primary goal of this sub-study was to investigate and refine criteria for developmental evaluation, multiple iterations of analysis and data collection also drove the evolution of our original questions into the criteria proposed above.

Validation

Longitudinal and iterative data collection and analysis created opportunities to validate our emerging interpretations through extended observation, triangulating among categories of

evidence, formal and informal member checking, resolving "negative cases," and mining the academic literature. Particularly important were our interviews with staff members in the New Tech hub, which provided opportunities to present, discuss, and refine the interpretations reported below.

Indeed, our data collection and analysis were intentionally structured to be consistent with the notion of developmental evaluation: that is, as a context within which we could function as "critical friends" who provided feedback and analysis over the course of the study. That experience, in turn, was instrumental in fostering the notion of actually developing a formal, theory-based method of developmental evaluation that could be enacted with reliability and validity in other school improvement networks. Over the course of our study, we developed relationships that allowed us to question and challenge hub and school staff, and that empowered hub and school staff to push back on our interpretations.

As discussed below, a final step will be to discuss the findings reported here with New Tech stakeholders in order to further refine our interpretations, to incorporate contrary interpretations, and to assess the usefulness of our analysis to these stakeholders in making strategic decisions about moving forward.

Findings

Our findings should not be read as *criticism* of the New Tech Network. Rather, they should be read as an *empathetic-yet-critical analysis* of the New Tech Network, given the circumstances that the network is facing at this point in its history and our interest in exploring the possibility of providing formative feedback for consideration by stakeholders in moving forward.

Our analysis suggests a core set of strengths in the New Tech Network, the product of ten years of hub-school collaboration: for example, a formal design for practice; formal routines and guidance for establishing organizational infrastructure in schools; formal principles to guide the enactment of instructional, leadership, and coaching practice; and a tradition of (and formal resources supporting) the social retention and reproduction of capabilities. These strengths are used to support schools in operationalizing principles of project-based learning to achieve school-determined learning outcomes. Our evidence suggests that this strategy is linked to New Tech's emergence in a newly-created school operating in supportive environments; to initial scale up in schools without histories of chronic underperformance; and to initially-small numbers of like-minded (and similarly-experienced) teachers, school leaders, and New Tech coaches.

However, our analysis also suggests that current conditions warrant evaluating the New Tech Network as an evolutionary enterprise: specifically, limits on social mechanisms for retaining and recreating capabilities resulting from a rapid increase in the installed base of schools; and general weaknesses in available knowledge to support the practice, leadership, and coaching of project-based learning. While doing so reveals important strengths (as described above), it also suggests a set of interdependent vulnerabilities: specially, the tradition of operating as a hybrid, shell-and-incubation enterprise in which schools work within a conventional organizational design to operationalize project-based learning to achieve school-determined goals; weaknesses in formal supports for base-level operations; weaknesses in formal supports for adaptive use (and for learning from adaptive use); and weaknesses in the infrastructure and capabilities of the hub to support evolutionary learning.

From the perspective of the evolutionary logic, the above-described weaknesses interact to inhibit the development of a formal knowledge base supporting the replication of capabilities for project-based learning across large numbers of high schools. That, in turn, places a premium on two key resources that are likely to grow weaker as New Tech recruits both more (and more variable) schools: social mechanisms for retaining and recreating capabilities for project-based learning; and prior capabilities to enact (and to learn from enacting) project-based learning. Our conjecture, again, is that such conditions predict a "Matthew effect" or "digital divide" likely to manifest as variable (or weak) implementation and effects in summative impact evaluations.

While our evidence suggests that understandings of these vulnerabilities are emerging among hub staff members, and while a collection of new initiatives are under way to address them, our analysis does not suggest that the New Tech hub is reconsidering its established shell-and-incubation strategy in favor of something akin to an evolutionary strategy. Moreover, our analysis suggests a set of key issues that complicate moving in that direction.

The net result has the New Tech Network poised between a promising past and an uncertain future, with important strategic decisions to be negotiated among its stakeholders. Such findings should not come as a surprise. Indeed, project-based learning is an approach to authentic instruction novel by the standards of mainstream US public education -- an approach that reformers since John Dewey have struggled to introduce and institutionalize in small numbers of schools, never mind a state-sized network of public high schools. That can be understood as New Tech's raison d'être, as well as its essential challenge: collaborating with schools to overcome knowledge deficits in developing and replicating capabilities for enacting an uncommon, difficult, but (in the view of many) compelling approach to instruction.

Conditions Warranting Evaluation as an Evolutionary Enterprise.

In our analysis, the goal of the New Tech Network to support the enactment of a school-wide design for project-based learning in a large and diverse group of high schools warrants analysis of the network as an evolutionary enterprise. This warrant rests, in part, on practical limitations and time demands in the social retention and reproduction of capabilities across a nation-wide, state-sized network of high schools. These include practical and logistical limits on mentorship and apprenticeship that would involve the exchange of practicing teachers, leaders, and New Tech coaches between schools, in all content areas, for extended periods of time; the fact that, at this time, nearly 70% of New Tech schools have three years or less experience with the network; and (owing to a proportional increase in the number of newly-hired New Tech coaches) reports of strains on social mechanisms for developing coaching capabilities.

This warrant also rests on equivocal evidence of available knowledge supporting New Tech's ambitions for schools. On the one hand, we identified formidable (and active) academic and professional literatures on project-based learning in K-12 education and beyond; established traditions of professional preparation and practice using comparable instructional methods (e.g., Montessori and Waldorf schools); organizations chartered to promoted understandings, practices, and research on project-based learning (e.g., the Buck Institute for Education); and summaries and meta-analyses - of varying rigor - asserting positive effects of project-based and problem-based learning on student achievement.

On the other hand, we did not identify anything like an integrated body of professional knowledge, practices, and education that would support school-wide enactment of interdisciplinary, project-based learning in a large and diverse network of high schools. For example, we identified two widely-cited reviews of research that raised concerns about the lack of a unified theory or model of project-based learned, along with concerns about the lack of

research on specific instructional methods, teacher-designed projects, and effectiveness with underperforming students (Thomas, 2000; Hmelo-Silver, 2004). Further, among the What Works Clearinghouse, the Best Evidence Encyclopedia, and a comprehensive review of research on comprehensive school reform (Borman et. al, 2003), we identified only one high-school level, whole school program that both used elements of project-based learning and was identified as having highly promising evidence of effectiveness (Expeditionary Learning Outward Bound). Finally, in an informal review of degree requirements in leading undergraduate and graduate education programs as identified by *US News and World Report*, we did not find any that focused specifically on project-based learning in the pre-service education of high school teachers, school leaders, or teacher educators.

The equivocality of (and potential weaknesses in) professional knowledge, practices, and education would appear to interact with legacy conditions in high schools and their environments both to explain the absence of school-wide, interdisciplinary project-based learning and to complicate its introduction. Such conditions include professional education that emphasizes disciplinary content over pedagogy; policies that require teachers to be highly qualified in specific content areas (and not to have interdisciplinary capabilities); labor market dynamics that result in weaker teachers working in high poverty schools; teachers working in isolation, in balkanized and discipline-specific departments; department chairs and school leaders struggling to develop the capabilities and political capital to engage in close-to-the-bone instructional improvement; and both legacy and standards-based environments that provide few incentives (and many disincentives) for teachers and students to enact novel forms of instruction in pursuit of "21st century skills" and "deeper learning."

Research on several leading education reform initiatives detail the knowledge demands of enacting novel-and-ambitious instruction at a large school: for example, Man: A Course of Study; the Coalition of Essential Schools; the Accelerated Schools Project; and America's Choice. In our review, the strongest evidence supporting our argument comes for a highly developed program of research and development from The University of Michigan and Northwestern University on the large-scale enactment of technology-mediated, project-based science in high poverty middle schools and high schools (Blumenfeld et. al, 2000; Krajcik & Blumenfeld, 2006).[22] Lead by leading scholars in the learning sciences, this initiative was cited in one review of research as one of three leading centers of project-based learning in the United States (Thomas, 2000). Though partial by the standards of the New Tech Network (e.g., one content area, three grade levels, 26 schools, and 63 teachers), this initiative still functions as a useful metric with which to consider both the work being undertaken in the New Tech Network and the standards to which it will likely to held.

The researchers reported that, through seven years of collaboration among project members and schools, they succeeded in developing five project-based science units that, with coaching, could be enacted successfully at a large scale (Krajcik & Blumenfeld, 2006). Researchers reported that their work was complicated by interdependent weaknesses and challenges among schools, their programs, their hub organizations, and their environments (Blumenfeld et. al, 2000). Moreover, they linked their success to an approach to design research that bears remarkable similarity to the evolutionary logic, with particular emphasis on working with schools iteratively and over time to produce, use, refine, and evaluate a set of highly-specified projects and associated professional development resources for use by (and with)

[22] See references lists from the two cited articles for a more complete list of publications from this initiative. See, also, http://hi-ce.org/.

teachers and students (Blumenfeld et. al; Krajcik & Blumenfeld, 2006). While they expressed concern that heavier-than-anticipated formalization of instructional resources made the program "somewhat closed" when compared to their original vision, they reported that as a necessary tradeoff for effective implementation and outcomes at scale (Krajcik & Blumenfeld:673).

Noteworthy is that, while the researchers published positive results in multiple peer reviewed journal articles, none of these reports met the evidence standards of the What Works Clearinghouse: a useful proxy for the standards to which the New Tech Network is likely t be held in summative impact evaluations.

Replication Infrastructure

By the evolutionary logic, if essential knowledge is either weak or non-existent, and if it is difficult to share knowledge person-to-person and organization-to-organization, then it becomes incumbent upon the hub both to produce essential knowledge and to devise other means of recreating it in outlets. That begins with a design for practice derived from a functioning template, as well a replication strategy, in which exploitation and exploration function as complementary learning processes that yield formal knowledge of where, what, and how to replicate.

Within the New Tech Network, despite a formal design for practice and established templates, our analysis suggests a disconnect between (a) the argued need to operate as an evolutionary enterprise and (b) an established practice of operating as a shell-and-incubation enterprise in which schools operationalizing this design for practice within a New Tech-specific organizational design . That, in turn, results in a general orientation away from strategies, values, and understandings central to developing a formal knowledge base supporting the replication of capabilities for project-based learning in large numbers of schools.

On the one hand, a key strength of the New Tech Network is that it has a formal design for practice, with proof of concept evidenced by both an historical template (Napa New Technology High School) and a series of "demonstration sites" that have met internal criteria for high-quality implementation and outcomes.[23] The design for practice centers on four key roles, the functional responsibilities of which are formalized primarily using principles of practice as detailed in rubrics and other documents.[24] *Teachers* are to collaborate in pairs to design, enact, and assess interdisciplinary project-based learning opportunities that incorporate technology in novel ways, respond to district and state standards, and leverage community partners. *Students* are to collaborate in small groups to engage academic content in the context of co-enacted projects, using information technology (rather than textbooks) as a primary resource and producing artifacts (rather than conventional assessments) as evidence of content mastery and skill development. Akin to principals, *directors* function as the primary on-site change agent, with responsibility for instructional organization and management, program administration, recruiting students and teachers, serving as the primary liaison with the hub, and maintaining relationships with community partners. *Advocates* function as supplemental, school-level leaders responsible for serving as liaisons to the hub and to the New Tech coach, organizing agendas for coaching days, and serving as teacher-leaders on project-based learning.

On the other hand, despite emerging conditions that (arguably) favor an evolutionary enterprise, the New Tech Network has, since its inception, been structured and operated as a

[23] We did not independently verify the performance level of the template or demonstration sites. Moreover, absent careful pre- and post-observations of each template, it is not clear whether these schools entered the New Tech Network with prior capabilities for project-based learning or if they developed those capabilities through participation in the network.

[24] These include the "Teacher Rubric", "Principal Evaluation Rubric", "School Success Rubric", and "TNT Site and Advocate Duties" document, all available to schools through Echo, the New Tech Network's online learning management system.

hybrid shell-and-incubation enterprise.[25] By way of a "shell", the strategy has the hub and schools collaborating to establish essential (and conventional) school-level infrastructure and environmental supports. By way of "incubation," schools are then responsible for working within that shell to leverage New Tech-provided principles, New Tech-provided coaching, and social relationships with other schools in operationalizing project-based learning in their own, local contexts and in pursuit of school-determined goals.[26] As explained by one New Tech staff member in a 2012 interview, "as a school development organization, we're trying not to talk about the single model but, rather, design principles, because we think that that's a better way of being more inclusive about what we do in co-designing the schools." Also in a 2012 interview, another staff member put it more plainly, explaining that the goal is "to use our design to support their vision."

As explained above, our evidence suggests that New Tech's shell-and-incubation strategy is more an artifact of interacting conditions during initial scale up and less a strategic decision anchored in careful analysis of social mechanisms or available knowledge. In interviews in the winter of 2012, New Tech staff members advanced three lines of argument supporting the strategy. The first centered on challenges to developing formal resources for use in all schools: for example, the financial and human resources needed to develop formal resources in all content areas; limitations in practical knowledge and experience among New Tech staff members (especially with respect to school-level leadership); and variation in school, districts, and state curriculum, standards, and assessments. The second was anchored in beliefs

[25] In our interviews in the winter of 2012, two hub staff members actually used the term "incubator" in describing the New Tech replication strategy.

[26] Examples of social relationships include initial visits to template schools to establish a vision for success, two days of "shadowing" opportunities for new leaders and staff prior to Year 1 implementation, regularly-scheduled "Meetings of the Minds" that bring together teachers from geographically-proximal schools, and ad hoc visits during the school year. It also includes the social exchange of classroom projects using Echo, the New Tech Networks online learning management system.

in both the value of and possibilities for school-based invention. As one New Tech staff member explained, "for schools to own it, they need to design it". The third centered on an aversion among New Tech staff members to formal, externally-developed routines and guidance for practice (and their perception that teachers and leaders share their view).[27] Rather than being understood as "enabling" (as in the evolutionary logic), such formalization was referred to skeptically (if not pejoratively) by some New Tech staff members as a "cookie cutter" and "plug and play" approach.

Formal Supports for Base-Level Operations

By the evolutionary logic, formal supports for base-level operations would include routines and guidance supporting the conventional, coordinated enactment of role-specific functional responsibilities in ways that yield acceptable outcomes, independent of adapting routines and guidance to address problems and improvement performance. Consistent with its history operating as a shell-and-incubation enterprise, and consistent with a general aversion to formal supports for practice, our analysis suggests weaknesses in formal supports for recreating conventional, coordinated base-level operations in schools. Absent a repurposing of New Tech's online learning management system, the result is the absence of an essential mechanism for exploiting knowledge at scale (e.g., projects designed and tested, problems solved, and lessons learned), as well as a mechanism for preventing enduring problems of early implementation in schools (e.g., non-implementation, cooptation, and/or regression to past practice).

As reported by long-serving New Tech staff members, a decade of annual recruiting cycles has yielded a core strength of the networks: specifically, considerable knowledge of "where" to replicate, as captured in formal routines and guidance for identifying, recruiting, and

[27] Our evidence suggests that this view is not shared uniformly among the teachers, school leaders, or schools in our study.

securing commitments from schools and districts, with the goal of establishing initial conditions supporting implementation. As detailed in a formal planning timeline, this includes pre-adoption visits by school and district staff to a functioning template to establish a sense of new possibilities for their schools. It includes a formal assessment and approval process focused on evaluating essential school-level infrastructure as formalized in a "Conditions for Success" rubric: for example, school design and culture, instructional organization, technology, facilities, external partnerships, and staffing. The process culminates in a formal contract that details the relationship between the hub and schools. And it includes follow-up visits in which a "School Success Rubric" is used to evaluate success establishing and maintaining key organizational arrangements.

While New Tech's "Conditions for Success" rubric calls for school autonomy in developing staffing procedures that "reflect the specific requirements of the model," the "where" of establishing new schools does not include routines and guidance for assessing existing capabilities among teachers for enacting project-based learning, nor among school leaders for supporting project-based learning. Rather, our analysis suggests that the network is open to (and actively recruiting) schools with weak initial capabilities, as evidenced by its recent i3 initiatives and by efforts to recruit schools that (by virtue of their underperformance) are eligible for school improvement grants with which to fund their participation. Indeed, our observations suggest considerable variation in initial capabilities.[28]

[28] Available research would predict this outcome. For example, research on comprehensive school reform found that highly-formalized program adoption processes were often ineffective in establishing initial conditions supporting successful implementation, and that they resulted in more variable (and often weak) pools of newly-enlisted schools than expected. Complicating issues included newly-available school improvement funding that created incentives for less-than-earnest districts and schools to adopt programs; the fact that much of this funding targets underperforming districts and schools prone to cycles of adoption, weak or non-implementation, and subsequent abandonment; and program providers compromising their own selection processes both to increase their prominence and to secure fees from schools (Datnow, 2000; Peurach, 2011).

Absent assessments that ensure base-level capabilities upon entering the network, argued weaknesses in professional knowledge, education, and experience (combined with argued weaknesses in social supports) place a premium on providing newly-enlisted teachers with formal routines and guidance for enacting a version of authentic, collaborative instruction that likely deviates in fundamental ways from their past practice. It also places a premium on formal routines and guidance for supporting school leaders and New Tech coaches as they, in turn, support teachers (a form of practice that is likely new to them, as well).

As described above, New Tech's design for practice does include role-specific and school-wide rubrics that formally specify base levels of performance. For example, the Teacher Rubric describes "proficient" teachers as (among other things) effectively managing groups; differentiating instruction; remediating within the context of a project; building skills for student collaboration; helping students to use performance rubrics to guide their collaborative work; regularly assessing learning outcomes; and using *Echo*, New Tech's online learning management system. As another example, the Principal Evaluation Rubric describes "proficient" leaders as (among other things) establishing a clear vision and mission; modeling cultural expectations (including open communication and constructive feedback); providing opportunities to teachers for improvement and support; developing and nurturing external relationships; and meeting district and legal requirements.

However, consistent with New Tech's established shell-and-incubation strategy, the "what" and "how" of establishing conventional, coordinated base-level operations are not formalized in coordinated routines and guidance. For example, the hub organization does not provide new teachers with a set of detailed, tested, and refined projects that teachers can use to develop common practices for (among other things) managing groups, differentiating

instruction, remediating in response to problems, building students' collaboration skills, and assessing learning outcomes in a project-based classroom. Further, neither new school leaders nor incoming New Tech coaches are provided with a curriculum to support teachers in establishing "proficient" capabilities for enacting project-based learning, nor with formal, coordinated routines for observing instruction, evaluating performance and outcomes, debriefing observations, and providing practice-based support for teachers (never mind the differentiation of such resources by individual content areas, interdisciplinary pairs of content areas, grade level, or teachers' capabilities). Finally, teachers, school leaders, and New Tech coaches are not provided with formal guidance that filters, synthesizes, and integrates the available literature into a coherent set of base-level understandings to be shared across all New Tech schools: for example, as related to the difficult history of project-based learning in US public education; its theoretical foundations; its manifestations in particular content areas; its challenges for students; or other knowledge and information to support specific practices.

The closest approximation to a source of readily-useable routines and guidance is *Echo*, which functions as a mechanism through which teachers can publish and share self-designed projects. However, in terms of routines, New Tech staff reported that *Echo* was not designed to provide new teachers with readily-useable projects but, instead, to provide them with models of well-designed projects that they could either emulate or adapt in devising their own projects. Even then, New Tech staff reported that weak vetting procedures resulted in the incorporation of many weak and questionable projects. Further, in terms of guidance, while *Echo* contains an extensive library of New Tech-provided materials with information germane to the program, this guidance is not synthesized and integrated in order to establish common, base-level understandings of the program, its history, intended practices, and professional knowledge

across all New Tech schools. Instead, our analysis is that this guidance mirrors the broader literature on project-based learning: sprawling and lacking coherence. Indeed, in the schools that we observed, teachers and school leaders uniformly expressed frustration with *Echo* as a source of useable projects and guidance.

Formal Supports for Adaptive Use

By the evolutionary logic, adaptive use in schools is critical both to addressing local exigencies and introducing new knowledge that supports network-wide improvement. Indeed, in terms of adaptive use, the essential "what" to be replicated across New Tech schools are capabilities for designing projects for use in classrooms. Absent New Tech-provided projects, teachers' work designing projects actually *becomes* base-level operations in New Tech schools. Further, as with base-level operations, there is no "where" mechanism to identify teachers ready for such work, nor for identifying teachers ready and able to *learn from the experience* of enacting the resulting projects -- capabilities that cannot be assumed, especially in underperforming schools. Rather, this work begins immediately, school-wide, during summer training prior to Day 1 of Year 1 as a New Tech school, and it continues thereafter.

This is the essential work of New Tech's shell-and-incubation strategy, and there is no overstating its complexity: school-determined interdisciplinary teacher teams at all grade levels; designing a form of authentic instruction for which they may have no prior experience or training; supported by equally-novice school leaders and by variably experienced coaches; never mind actually *enacting* the resulting projects successfully with students who, themselves, are likely unfamiliar with project-based learning and who may well lack capabilities for productive, technology-mediated, collaborative learning with peers.

Our analysis suggests that the New Tech Network provides comparatively stronger formal supports for adaptive use. Even so, our analysis also suggests key weaknesses: both in formal supports *for* adaptive use and in formal supports *for learning from* adaptive use. The result is much opportunity for exploratory learning among New Tech schools, though great risk of precisely the problem reported by New Tech staff members with respect to *Echo*: the accumulation of formalized "noise" throughout the network, rather than formalized practices and understandings with evidence of positive effects on student outcomes.

The New Tech Network provides a collection of formal resources detailing the "what" of teachers' design work, as well as the "how" of school leaders and New Tech coaches in supporting that work. Complemented by a strong emphasis on "fidelity to the model," these formal resources establish conventions for project-based learning within and between classrooms and schools (and, with that, prospects for replicable practice). Consistent with support for base-level operations, the principles of teacher-based design (and of supporting teacher-based design) are formalized in New Tech's many rubrics, with a focus on such issues as the comprehensiveness, rigor, and relevance of projects; the number of state standards that they address; the integration of content from other subject areas; and the use of real-world scenarios. Different from support for base-level operations, these principles are complemented by formal routines for use jointly by teachers, school leaders, and New Tech coaches. These routines include a "Project Planning Form" to guide the process of designing a project from scratch; *Echo*, which provides routines for organizing newly-created projects and associated materials, using them in the classroom, and sharing them throughout the network; and "Critical Friends Protocols" that structure collaborative assessment, evaluation, and reflection. Formal routines for coaches go further, to include strategies for "cognitive coaching."

Even with these comparative strengths, our analysis suggests key shortcomings in formal routines and guidance supporting the work of designing projects. For example, as with establishing base-level capabilities, New Tech does not include a formal curriculum for use by school leaders and New Tech coaches to scaffold teachers' efforts in project design, nor does it include a developmental process through which leaders and coaches could scaffold teachers from designing *selected components* of otherwise-specified projects to designing *entire projects* on their own. Again, teachers begin designing entire projects immediately, in the summer preceding Day 1 of Year 1 as a New Tech school, with novice leaders and variably-experienced coaches supporting them using the above-described routines. Further, as with base-level operations, this work is not supported by coherent, integrated guidance, such that design activity within and between schools is anchored in a set of common, core understandings: for example, about the development of an essential, driving question for projects in particular content areas and grade levels; the design of authentic, real-world tasks; the creation of developmentally-appropriate evaluation rubrics; or the meaning, substance, and coordination among district and state standards and assessments as they bear on project design. While *Echo* does provide access to documents addressing many of these issues, these resources, again, lack synthesis, integration, and coherence.

Our analysis also suggests key weaknesses in formal supports for *learning from* the work of designing and enacting projects so that projects evolve over time from "something that conforms to New Tech-specified conventions" to "something that shows evidence of working to improve student achievement." For example, New Tech does not structure a process for reviewing and revising teacher-designed projects in response to implementation or outcome evaluations, nor does it structure a process for the repetitive enactment of projects within or

between schools in order to revise them through experience and evaluation.[29] Rather than providing formal routines or resources to assess student achievement in specific academic content areas, the hub organization asserts that "currently, a single standard of measurement does not exist that can assess our vision of achievement. Most standardized tests simply do not measure critical thinking, collaboration, creativity, and communication skills" (New Tech Network, 2012g). Instead, student assessments and academic goals are teacher-determined and school-determined. Further, unless drawn from *Echo* and used without modification in another school (by all reports, a rare case), our analysis suggests that repetitive use of a project is limited to teacher teams deciding to re-use their own, self-designed projects, such that a second attempt at enacting the same project will not occur until a year after the first attempt -- if at all.

To be clear, New Tech's many rubrics include principles advocating for teachers, school leaders, and coaches to reflect on and improve the design and enactment of projects. Further, in our observation, New Tech coaches do support teachers and leaders in learning how to use Critical Friends Protocols for collaborative assessment, evaluation, and reflection. However, in our analysis, such work is weakly supported by coordinated routines and guidance, and such work is not formally structured into the school day in New Tech schools. Rather, such work occurs primarily in the context of school visits by New Tech coaches (roughly seven days per school year) or, alternatively, upon the initiative of individual school leaders.

Hub Infrastructure and Capabilities for Evolutionary Learning

By the evolutionary logic, leveraging network-wide activity in the service of building a formal knowledge base depends on infrastructure and capabilities in the hub to support collaborative, evolutionary learning. This infrastructure and these capabilities would enable the

[29] Such analyses would surely be complicated by the fact that, in most cases, the initial design of projects is typically confounded with their initial enactment in classrooms, making it hard to discern whether project design, project enactment, or some combination function to explain instructional outcomes.

hub to monitor local design and adaptation; evaluate and select program improvements based on their effectiveness; square them with changing environments; formalize them as routines and guidance in material or digital resources; and disseminate them back through the installed based.

The New Tech Network has elements of such an infrastructure, and this infrastructure has contributed to collaborative learning within the network. Even so, the infrastructure and capabilities for evolutionary learning in the New Tech hub are weak, largely as an artifact of its established shell-and-incubation strategy. Simply put, the neither the hub nor the network were conceptualized or structured to operate in this way. Moreover, despite some evidence of the hub moving in this direction, our analysis suggests that "evolving to evolve" will not be so straightforward.

Regarding its infrastructure and capabilities for learning, the strengths of the New Tech hub are decidedly social and decidedly "networky," with communication infrastructure and information processing capabilities anchored in relationships among hub staff, school staff, and others. Key resources include shared language; social relationships between New Tech coaches, leaders and teachers; occasional meetings and conferences (local and national) among both schools and New Tech staff; advisory groups of high-performing teachers and school leaders; and formally-identified "demonstration sites" that function as key thought partners. Further, hub staff report relationships with individuals and agencies beyond the network that have potential to function as key partners in improvement: for example, the KnowledgeWorks Foundation; participants in the Hewlett Foundation "Deeper Learning" initiative; several universities; and established reformers and reform initiatives. Finally, staff members report a culture of learning, innovation, and healthy competition among New Tech staff members that supports the open

exchange and debate of information and that drives a press for continuous improvement. Indeed, *Echo*, New Tech's most highly developed formal resource, is actually intended to support social interactions throughout the network.

As observed and confirmed with New Tech staff members, these strengths have interacted to support ongoing improvements in the New Tech Network. Some have been aimed at improving implementation and outcomes in schools: for example, the evolution of procedures for recruiting schools and establishing a conventional organizational "shell"; the evolution of *Echo* as a learning management system; movement to "problem-based" (rather than project-based) learning in mathematics; and the development and refinement of rubrics elaborating principles of practice in schools. Others have been aimed at increasing the legitimacy of the network, improving its standing in the reform community, and increasing its prospects for long-term sustainability: for example, efforts to secure philanthropic and government contributions; the rapid scale up of the network; and the switch to a fee-for-service financial strategy.

Even with these strengths, our analysis suggests the lack of a robust infrastructure supporting the development of a formal knowledge base. As discussed above, the New Tech Network lacks two key mechanisms supporting evolutionary learning. The first is a strategy of using formal resources to support the design and improvement of practices and understandings in schools (the *exploration* that drives evolutionary learning). The second is a strategy of using formal resources to establish conventional, base-level operations within and between schools (a mechanism that drives the *exploitation* of knowledge as it develops and matures within the network).

The weaknesses run still deeper. For example, over its history, the hub has not developed formal measurement tools, systems, or capabilities for reporting and rigorously

analyzing implementation and outcomes (of the program as a whole or its components), in part because such capabilities are not part of the professional preparation and experience of most long-serving staff members.[30] Further, while the hub has procedures in place to collect common performance indicators across schools (e.g., content mastery on state accountability assessments; AYP status; graduation rates; ACT and SAT scores; college course credits earned; college application and acceptance rates; and post secondary enrollment), hub staff reported problems both in collecting and analyzing these data: for example, schools not reporting their data; school-within-a-school organizational arrangements that complicate disaggregating data about New Tech students; and variation among state accountability assessments. Further, they also report having little data that would support rigorous comparisons with non-New Tech schools. Finally, beyond developing and maintaining *Echo*, the hub does not have highly developed capabilities for formalizing practice and understandings as routines and guidance.

Potential Implications for Implementation, Outcomes, and Impact Evaluation

In our analysis, weak infrastructure and capabilities in the hub for developing a formal knowledge base push precisely on weaknesses and risks that are emerging in the New Tech Network as a consequence of its rapid growth: social mechanisms for retaining and recreating knowledge; and the prior capabilities of teachers, school leaders, and New Tech coaches to enact and support project-based learning. We observed that social mechanisms are further weakened by New Tech's dependence on teacher-designed (vs. hub-designed) projects, which

[30] By contrast, consider the above-cited program of research on project-based learning in middle school science, which was fielded by researchers in the learning sciences with highly-developed capabilities for small-scale experiments. Consider, also, Success for All, a comprehensive school reform program founded by psychologists at Johns Hopkins University with highly-developed capabilities both for small-scale experiments and for large-scale evaluation.

result in instructional practice sufficiently particular as to complicate deep collaboration across schools.[31]

Dependence on social mechanisms and prior capabilities, in turn, risks a "Matthew effect" or digital divide likely to manifest as variable (if not weak) implementation and outcomes on summative impact evaluations. One conjecture it that, the stronger a school's social network, its prior capabilities for project-based learning, and its capabilities to learn from experience, the more likely it will be to use opportunities, relationships, and resources afforded by the New Tech Network to enact interdisciplinary project-based learning in ways that lead to improved student outcomes. The corollary conjecture is that the weaker a school's social network, prior capabilities for practice, and prior capabilities to learn from practice, the less likely it would be to experience successful implementation and outcomes, and the more likely it would be experience enduring problems of US education reform: non-implementation; rote compliance; cooptation; regression to past practice; or some combination of these. While we do not have outcome measures, that is the broad pattern of implementation that we will report in forthcoming analyses of implementation drawn from broader study within which this developmental evaluation is situated.

Much of this is not news to New Tech staff members. Rather, at the same time that they reported success within the network, hub staff were frank in discussing how they leveraged the above-described learning infrastructure to identify problems of implementation and outcomes in the context of rapid scale up. For example, as described above, they reported the accumulation of many weak and problematic projects in *Echo*; the accumulation of a smaller number of well-designed projects; yet the accumulation of few (if any) "projects that work" as evidenced by

[31] This critique derives directly from the typology of teacher collaboration developed by Little (1990), as well as from our observations of collaboration among New Tech schools.

rigorous analysis of their effects on student outcomes. Further, they reported a weakening of social mechanisms for retaining and reproducing knowledge (even though two hub staff members objected to our interpretation of generally-weak knowledge in broader environments and increasingly-variable initial capabilities among newly-recruited schools). Finally, they reported a version of the "Matthew effect" or digital divide described above: teachers with weak initial capabilities adhering tightly to New Tech-supplied guidance for designing projects and feeling successful as a result, despite the absence of attention to the quality of implementation, the meaningfulness of students' experiences in the classroom, or the quality of student outcomes.

Emerging understandings in the hub, in turn, have fostered an agenda for improvement: for example, the recognized need to provide more formal support to new teachers, school leaders, and coaches; to develop systems and capabilities to support more rigorous evaluations of program effectiveness; and to develop capabilities in the hub to create new types of material and digital resources for use by teachers, school leaders, and trainers. In fact, the New Tech hub has secured funding, hired staff, and/or launched initiatives on each of these points: for example, eliminating low-quality projects from *Echo's* project library; experimenting with a set of conventional "starter projects" for use in new schools; creating and/or reviewing formal resources for new school leaders and New Tech coaches; developing its internal research capabilities; participating in efforts by the Hewlett Foundation to develop measures of "deeper learning"; incorporating frameworks for discussing the interplay between "technical" and "adaptive" change; and collaborating with schools to develop literacy and math modules that can be incorporated into teacher-design projects to address the Common Core State Standards.

Even with these initiatives, our analysis does not suggest a wholesale reconsideration by the New Tech hub of its established shell-and-incubation strategy, nor deliberate coordination of the above-described agenda and initiatives in moving toward an evolutionary strategy. Rather, the New Tech Network appears to be "evolving to evolve," by incrementally incorporating strategies and mechanisms that have potential to accumulate as the sort of learning infrastructure required of an evolutionary enterprise (though absent an explicit logic or strategy to guide movement in that direction).

That said, our analysis suggests that at least three issues that complicate continued movement toward an evolutionary enterprise. The first is that old habits die hard. Many New Tech staff members remain ideologically committed to its shell-and-incubation strategy and ideologically opposed to supporting practice using hub-formalized routines and guidance. Indeed, many New Tech staff members came up through the ranks as classroom teachers in New Tech schools and, thus, lack other perspectives from which to reflect critically on strategies and experiences in the New Tech Network.

The second is the real cost (in terms of human and financial resources) in developing the type of research and development capabilities needed to support an evolutionary strategy. Moreover, these resource demands are likely to be experienced just as a large number of New Tech schools transition out of their initial contract, resulting in a reduction in fees from schools or, possibly, a reduction in the installed base of schools (depending on whether schools elect to continue or to drop out). Indeed, given the recruitment of increasing numbers of initially underperforming schools, and given histories of program adoption-and-abandonment in many such schools, the potential loss of schools is very real.

The third is that two new initiatives risk increasing the knowledge demands on the network, drawing attention and resources away from the high school model, and introducing new uncertainty. The first is an effort to experiment with school-wide, project based learning in middle schools to establish a "feeder pattern" into high schools. The second is an effort to experiment with district collaboration to establish a more stable environment in which both middle schools and high schools can operate. Indeed, in moving in this direction, it is not clear that members of the hub organization understand these moves as placing new knowledge demands on the enterprise. In contrast to the above-argued deficits in available knowledge, one New Tech staff member explained in a 2012 interview: "I think proof of concept in developing a middle school model is going to be fairly easy to do, to develop projects that are developmentally appropriate. I'm confident that's not going to be a big step for us."

Topics for Formative Conversation among Stakeholders

The preceding analysis stands as evidence of the potential power of using our proposed indicators to structure a developmental evaluation of a school improvement network. In contrast to the more typical focus on the effectiveness of program implementation and outcomes, these criteria focus attention keenly on the *logical antecedents* to effective implementation and outcomes: specifically, matters related to building and continuously improving a formal knowledge base supporting the large-scale replication of capabilities. Rather than criticism, the intent is to provide funders, reformers, and schools with empathetic-yet-critical perspective on the strengths and vulnerabilities of their network in an evaluation climate increasingly pressing for evidence of replicable effectiveness.

All told, the experiences of the New Tech Network appear to be par-for-the-course when it comes to establishing a large-scale school improvement network: the rapid scale up of a

promising, partial, and problematic model; anchored in highly institutionalized replication strategies; with challenges arising from interactions among schools, the program, the hub, and environments; and with the hub collaborating with schools and others to "learn its way through," relying primarily on social (rather than formal) mechanisms for retaining and recreating knowledge. In the short term, prospects for favorable summative impact evaluation appear uncertain. Additional time, patience, and resources appear to be required in order to expand and refine the formal knowledge base supporting the scale up of the New Tech Network.

Assuming additional time, patience, and resources, the real test of the value of this developmental evaluation lies in its usefulness in fostering new types of formative discussion among stakeholders in the New Tech Network. To the extent that they are open it, we propose four topics for formative conversation:

1. Strengths of the network: We would encourage a comprehensive, collaborative review of strengths that have accumulated throughout the New Tech Network over its history. Strengths mark ground gained and returns on investment, and they are a foundation on which to continue building. Reviewing the preceding analysis, our list of strengths includes the establishment of a practice-focused school improvement network; a formal design for practice; routines and guidance for establishing organizational infrastructure in schools; formal routines for supporting initial project design; knowledge retained socially in the original template, demonstration sites, and communities of experienced coaches; *Echo* as a potentially-powerful resource for retaining and communicating knowledge and information; and a formidable social infrastructure for collaborative learning.

2. Premises of developmental evaluation: Our argument for developmental evaluation rests on four premises that may or may not be understood, shared, or valued among stakeholders: prospects for summative impact evaluation; uncertainty in the work of school improvement networks; capabilities as a prerequisite to effective implementation and outcomes; and collaborative, evolutionary learning as a means of building a formal knowledge base supporting the large-scale replication of capabilities.

3. Conditions supporting evaluation as an evolutionary enterprise: Our analysis of the New Tech Network is predicated on an analysis of social mechanisms and existing knowledge supporting the large-scale replication of project-based learning. Since stakeholders, again, may or may not share or value our analysis, it also merits collective, critical consideration. One issue is whether our analysis is on the mark. Another issue is to consider possible decisions that could alter these conditions in ways that would support continuing to operate as a shell-and-incubation enterprise: for example, reducing the size of the network; limiting the pace of growth; working in a subset of content areas particularly amenable to project-based learning; identifying, vetting, and incorporating commercial or other curricula designed to support project-based learning; and enlisting only teachers, leaders, and schools with prior knowledge and experience with project-based learning.[32]

4. Movement toward an evolutionary strategy: If stakeholders agree that our premises and analysis of conditions are on the mark, then a next step is to collectively consider the strengths and weaknesses of the remainder of our analysis, and -- to the extent that it passes muster -- the possible implications for moving forward. Regarding replication infrastructure,

[32] Any such discussion should be balanced against a careful review of the above-cited research on project-based learning in middle school science to understand the conditions that ultimately warranted an evolutionary strategy. Further, any such discussion should be balanced against reports that limits on social mechanisms for retaining and recreating knowledge were reported to have been experienced immediately in the New Tech Network, in its template site (Borja, 2002).

this would include conversations about realigning resources, beliefs, and values to support an evolutionary strategy. Regarding base-level operations, this would include conversations about developing detailed projects and coordinated guidance to support novice teachers, coordinated with formal routines and guidance to support novice school leaders and trainers. Regarding adaptive use, this would include conversations about a developmental progression into designing full projects; routines and guidance to support more rigorous school-level evaluation of effectiveness; and creating opportunities for the repetitive enactment of projects. Regarding hub infrastructure and capabilities, this would involve conversations about activities to undertake (e.g., build research and development capabilities; develop deeper understanding of research on large-scale educational reform) and, possibly, not undertake (e.g., launch initiatives that increase knowledge demands on the network and that draw resources and attention away from the high school program).

Discussion

This analysis was motivated by our concern with agenda instability as a consequence of predictably equivocal (if not weak) summative impact evaluations of school improvement networks. The purpose of this analysis was to propose and investigate criteria for a new type of formative, developmental evaluation that would provide stakeholders with essential feedback in advance of impact evaluations, with the twin goals of both (a) creating the time needed to continue working and (b) improving prospects for favorable outcomes.

Toward that end, we critically analyzed conventional goals and processes of impact evaluations from the perspective of research on school improvement networks. Further, we proposed a logic and complementary criteria for developmental evaluation, anchored in theories of evolutionary economics and in research on organizational replication in both the commercial

and education sectors. Finally, we investigated those criteria in a developmental evaluation of a leading high school-level school improvement network to demonstrate their usefulness in the empathetic-yet-critical analysis of logical antecedents to successful impact evaluation.

There are limitations to the analysis, to be sure. Some of these limitations are in our logic and methods of developmental evaluation. For example, both the evolutionary logic and our proposed criteria are still nascent, and they are sure to evolve with subsequent attempts at developmental evaluation. This is but one theoretical perspective in which to anchor developmental evaluation, and an emergent one at that. Further, the research methods used for this developmental evaluation (longitudinal, embedded, ethnographic case study) are costly in terms of time and human resources, slow by the standards of the information needs of stakeholders, and hardly the type that could be enacted in all but a small subset of school improvement networks

Some of these limitations are tied more directly to our developmental evaluation of the New Tech Network. For example, the preceding is an initial developmental evaluation using our proposed criteria. Complementary analyses are needed to critically examine such issues as the content of routines and guidance; variation in use of these resources within and between schools, districts, and states; and the work of hub organizations in leveraging adaptive use in schools as resources for network-wide improvement.[33] Further, as discussed immediately above, our analysis stops short of examining its usefulness to New Tech stakeholders. Finally, as reported above, our broader study was neither conceptualized nor designed specifically to support developmental evaluation. Rather, our recognition of the need for developmental evaluation actually emerged in the course of our broader study. As such, we did not use our

[33] For example, since our data collection was centered primarily in one state, a useful follow-up to this study would be to review its primary findings with stakeholders in others states in order to solicit additional (and possibly contrary) perspectives.

proposed criteria to structure data collection, possibly resulting in our failing to identify (among other things) formal resources that might challenge our interpretations.[34]

Even with those limitations, we argue that our analysis yields a rationale, logic, criteria, and supporting evidence sufficiently robust as to warrant further investment in pursuing developmental evaluation as a complement to impact evaluation. This warrant is further supported by recognition of the importance of school improvement networks to the national education reform agenda; the amount of money invested in them; and the information needs of the many funders, reformers, and practitioners vested in them.

To begin, a first step would be to turn the analysis back onto itself in developing a replicable method of developmental evaluation for use by external evaluators and (possibly) network stakeholders. Such a method would require routines, procedures, and associated tools to ensure the efficiency, validity, and reliability of developmental evaluations: for example, conventions for study design; standards of evidence; procedures for data collection; methods and standards of analysis; and standards and conventions for reporting. Even more so, it would require extensive, coordinated, yet manageable guidance to support the enactment of these routines and the interpretation of evidence: for example, broad-based historical knowledge on educational reform; more focused knowledge of school improvement networks and of the challenges of improving practice; disciplinary knowledge on the production and use of knowledge; and more.

A complementary step would be to develop replicable methods to support evaluators and stakeholders in making productive use of developmental evaluations. This, again, would require

[34] For example, we recognize that our analysis pays little attention to formal resources supporting the enactment of the student role in instruction -- something very important to successful project-based learning, and something often very difficult for students accustomed to more traditional forms of instruction. The existence of such resources could mitigate our interpretations, while their absence could amplify our interpretations.

routines and procedures: for example, a structured process that combines presenting findings, jointly interpreting them, and arriving at consensus for moving forward; protocols to structure new types of discussions around unfamiliar issues; and methods for airing interpretations and resolving disagreements. And this, again, would require extensive guidance, especially about the normative dimensions of the evolutionary logic of replication. Despite having a sound theoretical and empirical basis, the logic is anchored in positive, mutually-reinforcing synergies widely understood as logical opposites: for example, diffusion *and* incubation as interdependent strategies for scaling up; fidelity *and* adaptation as primary goals of implementation; and exploitation *and* exploration as complementary learning strategies.

Stepping beyond this particular analysis, yet another step would be to consider the possibility of alternative methods of developmental evaluation. For example, it seems both plausible and prudent to conduct a complementary analysis structured around the types of "network improvement communities" being supported by the Carnegie Foundation for the Advancement of Teaching (Bryk, Gomez, and Grunow, 2010). It seems equally plausible and prudent to match *network-focused* developmental evaluation with *environment-focused* developmental evaluation. For example, in research on comprehensive school reform, Glazer and Peurach (2012) found that the development of school improvement networks depends heavily on the emergence of a supporting "community infrastructure" that includes institutional supports, resources endowments, proprietary activity, and market functions. While our first proposed criterion includes elements of such an analysis, more thorough analysis of the community infrastructure has potential to provider stakeholders with strategic information every bit as valuable as formative feedback on their own networks.

One contrast to the proposed next steps would be to *not* act on the argued and demonstrated value of developmental evaluation: that is, to stay the course; generate predictably equivocal (and likely weak) impact evaluations; fan the rhetorical flames; withdraw support (either for specific initiatives or for the agenda as a whole); and start over. That strikes us as a gross violation of what Elmore (2004) describes as the "reciprocity of accountability," which holds that expectations for performance must be matched with commensurate efforts to develop the capabilities needed to realize that level of performance. It also strikes us as inefficient, owing to the loss of intellectual capital that accumulates in school improvement networks as a consequence of collaborative, experiential learning among hubs and schools, as retained both socially (e.g., in communities of practice) and formally (e.g., in routines, guidance, tools, and artifacts).

Conclusion

With high school reform occupying a prominent place on the national reform agenda, and with school improvement networks a leading strategy for large-scale reform, the time is right to incorporate formative, developmental evaluation as a complement to summative, impact evaluation in collectively considering the progress of school improvement networks. Doing so would require researchers willing to take up the cause, funders willing to support it, and school improvement networks willing to participate. Moreover, it would require creating political cover for hub organizations willing to open their historically-private work to a new and uncertain form of evaluation. Even so, the payoff could be formidable: improved returns on billions of dollars in public and private investment, certainly; but, more importantly, improved educational experiences and outcomes for millions of students otherwise underserved, both by their public schools and by the reformers charged with improving those schools.

References

Adler, P. S. & Borys, B. (1996). Two types of bureaucracy: Enabling and coercive. *Administrative Science Quarterly, 41 (1)*, 61-89.

Arrow, K. J. (1962). Economic welfare and the allocation of resources for invention. In R. R. Nelson (Ed.), *The rate and direction of inventive activity* (pp. 609-625). Princeton, NJ: Princeton University Press.

Arrow, K. J. (1974). *The limits of organization*. New York, NY: W. W. Norton and Company.

Baden-Fuller, C. & Winter, S. G. (2005). *Replicating organizational knowledge: Principles or templates?* Jena, Germany: Max Planck Institute of Economics, Papers on Economics and Evolution.

Berends, M., Bodilly, S. J., & Kirby, S. N. (2002). *Facing the challenges of whole school reform: New American Schools after a decade*. Santa Monica, CA: Rand.

Berman, P. & McLaughlin, M.W. (1975). *Federal programs supporting educational change, Vol. 4: The findings in review*. Santa Monica, CA: Rand.

Berman, P. & McLaughlin, M.W. (1978). *Federal programs supporting educational change Vol. 8: Implementing and sustaining innovations*. Santa Monica, CA: Rand.

Blumenfeld, P., Fishman, B. J., Krajcik, J. S., Marx, R. W., & Soloway, E. (2000). *Creating usable innovations in systemic reform: Scaling up technology-embedded project-based science in urban schools*. Educational Psychologist, 35 (3), 149-164).

Bodilly, S. J. (1996). *Lessons from New American Schools Development Corporation's demonstration phase: Prospects for bringing designs to multiple schools*. Santa Monica, CA: Rand.

Borja, R. R. (2002, May 29). Improving the practice of large-scale school reform. *Education Week, 21 (38)*, 26-31.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73 (2)*, 125-230.

Borman, G., Slavin, R.E., Cheung, A., Chamberlain, A., Madden, N.A., and Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal, 44 (3)*, 701-731.

Bradach, J. L. (1998). *Franchise organizations*. Boston, MA: Harvard Business School Press.

Bradach, J. L. (2003). Going to scale: The challenge of replicating social programs. *Stanford Social Innovation Review, Spring 2003*, 19-25.

Brewer, J. D. (2004). Ethnography. In C. Cassell & G. Symon (Eds.), *Essential guide to qualitative methods in organizational research*. Thousand Oaks, CA: Sage.

Brown, J. S. & Duguid, P. (1998). Organizing knowledge. *California Management Review, 40 (3)*, 90-111.

Buck Institute for Education. (2012). *What is PBL?* Retrieved June 1, 2012, from http://www.bie.org/about/what_is_pbl/.

Bulkley, K. E. and Burch, P. (2011). The changing nature of private engagement in public education: For-profit and nonprofit organizations and educational reform. *Peabody Journal of Education, 86 (3)*, 236-251.

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

Camburn, E., Rowan, B., & Taylor, J. T. (2003). Distributed leadership in schools: The case of elementary schools adopting comprehensive school reform models. *Educational Evaluation and Policy Analysis, 25 (4)*, 347-373.

Cameron, K. S. and Spreitzer, G. M. (Eds.) (2011). *The Oxford handbook of positive organizational scholarship*. New York, NY: Oxford University Press.

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher, 32 (6)*, 3-12.

Cohen, D. K., Moffitt, S. L. & Goldin, S. (2007). Policy and practice: The dilemma. *American Journal of Education, 113 (4)*, 515-548.

Cohen, D. K. & Moffitt, S. L. (2009). *The ordeal of equality: Did federal regulation fix the schools?* Cambridge, MA: Harvard University Press.

Cohen, D. K., Peurach, D. J., Glazer, J. L., Gates, K. G., & Goldin, S. (In press). *Improvement by design*. Chicago, IL: University of Chicago Press.

Cohen, D. K. & Spillane, J. P. (1991). Policy and practice: The relations between governance and instruction. In S. H. Fuhrman (Ed.), *Designing coherent education policy: Improving the system* (pp. 35–95). San Francisco: Jossey-Bass.

Cohen, W. M. & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly, 35,* 128–152.

Constas, M. A. & Brown, K. L. (2007). Toward a program of research on scale-up: Analytical requirements and theoretical possibilities. In B. Schneider & S. K. McDonald (Eds.), *Scale up in education: Volume 1 -- Ideas in principle* (pp. 247-258). Lanham, MD: Rowman & Littlefield Publishers, Inc.

Clyburn, G. (2011). *R&D ruminations: A hub by any other name*. Stanford, CA: Carnegie Foundation for the Advancement of Teaching.

Datnow, A. (2000). Power and politics in the adoption of school reform models. *Educational Evaluation and Policy Analysis, 22 (4)*, 357-374.

Datnow, A., Hubbard, L, & Mehan, H. (2002). *Extending educational reform: From one school to many*. New York, NY: Routledge Publishers.

Datnow, A., & Park, V. (2009). Towards the co-construction of educational policy: Large-scale reform in an era of complexity. In D. Plank, B. Schneider, & G. Sykes (Eds.), *Handbook of Education Policy Research* (pp. 348-361). New York, NY: Routledge Publishers.

DiMaggio, P.J. & Powell, W.W. (1991). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. In W.W. Powell & P.J. DiMaggio (Eds.), *The New Institutionalism in Organizational Analysis* (pp. 63-82). Chicago: The University of Chicago Press. (Reprinted from *American Sociological Review, 48*, 1983, pp. 157-160.)

Dosi, G., Nelson, R. R., & Winter, S. G. (2001). Introduction: The nature and dynamics of organizational capabilities. In G. Dosi, R. R. Nelson, & S. G. Winter (Eds.), *The nature and dynamics of organizational capabilities* (pp. 51-68). New York, NY: Oxford University Press.

Dutton, J. E., Quinn, R. E., and Cameron, K. S. (Eds.) (2003). Positive organizational scholarship: Foundations of a new discipline. San Francisco, CA: Berrett-Koehler Publishers.

Easterly, W. (2009). *The civil war in development economics*. Retrieved April 4, 2012, from http://aidwatchers.com/2009/12/the-civil-war-in-development-economics/.

Eisenhardt, K. M. & Martin, J. A. (2000). Dynamic capabilities: What are they? *Strategic Management Journal, 21 (10/11)*, 1105-1121.

Elmore, R. F. (2004). *School reform from the inside out: Policy, practice, and performance*. Cambridge, MA: Harvard Education Press.

Farrell, C., Nayfack, M. B., Smith, J., Wohlstetter, P., & Wong, A. (2009). *Scaling up charter management organizations: Eight key lessons for success*. Los Angeles, CA: University of Southern California, Center on Educational Governance, Rossier School of Education.

Feldman, M. S. & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly, 48*, 94-118.

Fine, G. A., Morrill, C., & Surianarain, S. (2008). Ethnography in Organizational Settings. In D. Buchanan and A. Bryman (Eds.), *The Sage Handbook of Organizational Research Methods*. Thousand Oaks: Sage.

Firestone, W.A. & Corbett, H.D. (1988). Planned organizational change. In N.J. Boyan (Ed.), *Handbook of research on educational administration* (pp. 321-340). New York, NY: Longman.

Foray, D., Murnane, R., & Nelson, R. (2007). Randomized trials of educational and medical practices: Strengths and limitations. *Economics of Innovation and New Technology, 16 (5)*, 303-306.

Furman Institute. (2012). *Investing in Innovation (i3): New Tech High Schools*. Retrieved May 6, 2012, from http://riley.furman.edu/education/projects/i3-new-tech-high-schools/investing-innovation-i3-new-tech-high-schools.

Glazer, J. L. & Peurach, D. P. (2012). School improvement networks as a strategy for large-scale education reform: The role of environments. *Educational Policy*. DOI: 10.1177/0895904811429283.

Glennan, T. K., Jr., Bodilly, S. J., Galegher, J. R., & Kerr, K. A. (2004). Summary: Toward a more systematic approach to expanding the reach of educational interventions. In T. K. Glennan, Jr., S. J. Bodilly, J. R. Galegher, & K. A. Kerr (Eds.), *Expanding the reach of educational reforms: Perspectives from leaders in the scale-up of educational interventions* (pp. 647-685). Santa Monica, CA: Rand.

Granger, R. C. (2011). The big why: A learning agenda for the scale-up movement. *Pathways, Winter* (28-32).

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal, 17 (Special Winter Issue)*, 109-122.

Hatch, T. (2000). What does it take to break the mold? Rhetoric and reality in New American Schools. *Teachers College Record, 102 (3)*, 561-589.

Hatch, T. & White, N. (2002). The raw materials of reform: Rethinking the knowledge of school improvement. *Journal of Educational Change, 3,* 117-134.

Hess, F. M. (1999). *Spinning wheels: The politics of urban school reform*. Washington, D.C.: The Brookings Institution.

Hmelo-Siler, C. E. (2004). *Problem-based learning: What and how do students learn? Educational Psychologist Review, 16 (3)*, 235-266).

Institute for Education Sciences. (2012a). *About IES: Connecting research, policy, and practice*. Retrieved March 9, 2012, from http://ies.ed.gov/aboutus/.

Institute for Education Sciences. (2012b). *Request for applications: Education research grants*. Retrieved March 9, 2012, from http://ies.ed.gov/funding/pdf/2012_84305A.pdf.

Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2010). *Handbook on impact evaluation: Quantitative methods and practice*. Washington, D.C.: The World Bank.

Krajcik, J. S. & Blumenfeld, P. (2006). Project-based learning. In R. K. Sawyer (ed.), *The Cambridge handbook of the learning sciences*. New York, New York: Cambridge University Press.

Lee, T.W. (1999). *Using qualitative methods in organizational research*. Thousand Oaks, CA: Sage.

Leithwood, K. & Menzies, T. (1998). Forms and effects of school-based management: A review. *Educational Policy, 12* (3), 325-346.

Little, J. W. (1990). The persistence of privacy: Autonomy and initiative in teachers' professional relations. *Teachers College Record, 91*, 509-536.

March, J. G. (1996). Exploration and exploitation in organizational learning. In M. D. Cohen & L. S. Sproull (Eds.), *Organizational learning* (pp. 101-123). Thousand Oaks, CA: Sage Publications. (Reprinted from *Organization Science, 2 (1)*, February, 1991).

McDonald, J. P., Klein, E. J., & Riordan, M. (2009). *Going to scale with new schools designs: Reinventing high schools*. New York, NY: Teachers College Press.

Meyer, J. W. & Rowan, B. (1978/1983). The structure of educational organizations. In J. W. Meyer and W.R. Scott (Eds.), *Organizational environments: Ritual and rationality* (pp. 71-98). Beverly Hills, CA: Sage Publications. (Reprinted from *Environments and Organizations*, pp. 78-109, by Marshall W. Meyer (Ed.), 1978, Jossey-Bass, Inc.).

Meyer, J. W., Scott, R. W., & Deal, T. E. (1983). Institutional and technical sources of organizational structure: Explaining the structure of educational organizations. In J. W. Meyer & W. R. Scott (Eds.), *Organizational environments: Ritual and rationality* (pp. 45–70). Beverly Hills, CA: Sage Publications. (Reprinted from Environments and Organizations, pp. 151–178, by H.D. Stein (Ed.), 1981, Philadelphia: Temple University Press.)

Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis*. Thousand Oaks, CA: Sage.

Mosteller, F. and Boruch, R. (Eds.) 2002. *Evidence matters: Randomized trails in education research*. Washington, D.C.: Brookings.

Muncey, D. E. & McQuillan, P. J. (1996). *Reform and resistance in schools and classrooms: An ethnographic view of the Coalition of Essential Schools*. New Haven, CT: Yale University Press.

National Governors Association et. al. (2008). *Benchmarking for success: Ensuring U.S. students receive a world class education*. Washington, D.C.: National Governors Association.

Nelson, R. R. & Winter, S. G. (1982). *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press.

New Tech Network. (2011). *New Tech Network: Services (version 10/2011)*. Napa, CA: New Tech Network.

New Tech Network. (2012a). *A vibrant network: Proudly supporting over 100 schools*. Retrieved April 02, 2012, from http://www.newtechnetwork.org/newtech_schools.

New Tech Network. (2012b). *NTN professional development: Year 1*. Napa, CA: New Tech Network.

New Tech Network. (2012c). *About us: What we do*. Retrieved April 02, 2012, from http://www.newtechnetwork.org/about_newtech.

New Tech Network. (2012d). Our model: What fuels our success. Retrieved April 02, 2012, from http://www.newtechnetwork.org/newtech_model.

New Tech Network. (2012e). New Tech Network Outcomes 2010-2011. Napa CA: New Tech Network.

New Tech Network. (2012f). Our team: Experts in school development and innovation. Retrieved April 02, 2012, from http://www.newtechnetwork.org/newtech_staff.

New Tech Network. (2012g). *NTN outcomes -- why we gather what we gather*. Retrieved May 20, 2012, from http://www.newtechnetwork.org/content/ntn-outcomes-why-we-gather-what-we-do.

Patton, M. Q. (2006). Evaluation for the way we work. *Nonprofit Quarterly (Spring)*, 28-33.

Patton, M. Q. (2011). *Developmental evaluation: Applying concepts to enhance innovation and use*. New York, NY: Guilford Press.

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

Peurach, D. J. (2011). *Seeing complexity in public education: Problems, possibilities, and Success for All*. New York, NY: Oxford University Press.

Peurach, D. J. (2012, February 29). Improving the practice of large-scale school reform. *Education Week, 31 (22)*, 20-21.

Peurach, D. J. and Glazer, J. L. (2012). Reconsidering replication: New perspectives on large-scale school improvement. *Journal of Educational Change, 13 (2),* 155-190.

Peurach, D. J., Glazer, J. L., and Lenhoff, S. W. (2012). Make or buy? That's really not the question. Considerations for systemic school improvement. *Phi Delta Kappan, 93 (7)*, 51-55.

Peurach, D. J. and Gumus, E. (2011). Executive leadership in school improvement networks: A conceptual framework and agenda for research. *Current Issues in Education, 14 (3)*.

Raudenbusch, S. W. (2007). Designing field trials of educational innovations. In B. Schneider & S. K. McDonald (Eds.), *Scale up in education: Volume II -- Issues in practice* (pp. 1-15). Lanham, MD: Rowman & Littlefield Publishers, Inc.

Rogers, E. M. (1995). *Diffusion of innovations*. Fourth edition. New York: The Free Press.

Rowan, B. (2002). The ecology of school improvement: Notes on the school improvement industry in the United States. *Journal of Educational Change, 3,* 283–314.

Rowan, B., Camburn, E., & Barnes, C. (2004). Benefiting from comprehensive school reform: A review of research on CSR implementation. In C. Cross (Ed.), *Putting the pieces together: Lessons from comprehensive school reform research* (pp. 1-52). Washington, D.C.: National Clearinghouse for Comprehensive School Reform.

Rowan, B., Correnti, R. J., Miller, R. J., & Camburn, E. M. (2009a). *School improvement by design: Lessons from a study of comprehensive school reform programs*. Philadelphia, PA: Consortium for Policy Research in Education.

Rowan, B., Correnti, R. J., Miller, R. J., & Camburn, E. M. (2009b). School improvement by design: Lessons from a study of comprehensive school reform programs. In G. Sykes, B. Schneider, & D. Plank (eds.) *AERA Handbook on Education Policy Research* (pp. 637-651). New York: Routledge.

Schneider, B. & McDonald, S. K. (2007a). Introduction. In B. Schneider & S. K. McDonald (Eds.), *Scale up in education: Volume 1 -- Ideas in principle* (pp. 1-15). Lanham, MD: Rowman & Littlefield Publishers, Inc.

Schneider, B. & McDonald, S. K. (Eds.) (2007b). *Scale up in education: Volume II -- Issues in practice*. Lanham, MD: Rowman & Littlefield Publishers, Inc.

Schochet, P. Z. (2008). *Technical methods report*: *Statistical power for regression Discontinuity designs in education evaluations* (NCEE 2008-4026). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Scholz, R. W. & Tietje, O. (2002). *Embedded case study methods: Integrating quantitative and qualitative Knowledge*. London: Sage.

Slavin, R.E. & Fashola, O.S. (1998). *Show me the evidence! Proven and promising programs for America's schools.* Thousand Oaks, CA: Corwin Press, Inc.

Smith, M. S. & O'Day, J. (1991). Systemic school reform. In S. H. Fuhrman & B. Malen, (Eds.), *The politics of curriculum and testing: The 1990 Yearbook of the Politics of Education Association* (pp. 233–267). New York: The Falmer Press.

Stebbins, R. A. (2001). *Exploratory research in the social sciences.* Thousand Oaks, CA: Sage.

Szulanski, G. & Winter, S. G. (2002). Getting it right the second time. *Harvard Business Review, 80 (3)*, 62-69.

Szulanski, G., Winter, S. G., Cappetta, R. & Van den Bulte, C. (2002). *Opening the black box of knowledge transfer: The role of replication accuracy*. Philadelphia, PA: University of Pennsylvania, Wharton School of Business.

Thomas, J. W. (2000). *A review of research on project-based learning*. Retrieved May 6, 2012, from http://www.bie.org.

U.S. Department of Education. (2010c). *Investing in Innovation Fund (i3) program: Guidance and frequently asked questions*. Retrieved 01/13/2011 from http://www2.ed.gov/programs/innovation/faqs.pdf.

Van de Ven. A. H., Polley, D. E., Garud, R., & Venkataraman, S. (1999). *The innovation journey*. Oxford: Oxford University Press.

Walsh, J. P. & Ungson, G. R. (1991). Organizational memory. *Academic of Management Review, 16 (1)*, 57-91.

Wernerfelt, B. (1995). The resource-based view of the firm: Ten years after. *Strategic Management Journal, 16 (3)*, 171-174.

Williams, A., Blank, R. K., Toye, C., & Petermann, A. (2007). *State education indicators with a focus on Title I 2002/2003*. Washington, D.C.: Council of Chief State School Officers.

Winter, S. G. (2003). Understanding dynamic capabilities. *Strategic Management Journal, 24*, pp. 991-995.

Winter, S. G. (2010). The replication perspective on productive knowledge. In H. Itami, K. Kusunoki, T. Numagami, & A. Takeishi (Eds.), *Dynamics of knowledge, corporate systems, and innovation* (pp. 85-124). New York, NY: Springer.

Winter, S. G. (2012). *Capabilities: Their origin and ancestry*. Philadelphia, PA: University of Pennsylvania, The Wharton School.

Winter, S. G. & Szulanski, G. (2001). Replication as strategy. *Organization Science, 12 (6),* 730-743.

Winter, S. G. & Szulanski, G. (2002). Replication of organizational routines: Conceptualizing the exploitation of knowledge assets. In C. W. Choo & N. Bontis (Eds.), *The strategic management of intellectual capital and organizational knowledge* (pp. 207-222). New York, NY: Oxford University Press.

Yin, R. K. (2009). *Case study research: Design and methods (fourth edition)*. Thousand Oaks, CA: Sage.

Zollo, M. & Winter, S. G. (2002). Deliberate learning and the evolution of dynamic capabilities. *Organization Science, 13 (3),* 339-351.