The Developmental Evaluation of School Improvement Networks

Donald J. Peurach¹, Joshua L. Glazer², and Sarah Winchell Lenhoff³

Abstract
The national education reform agenda has rapidly expanded to include attention to continuous improvement research in education. The purpose of this analysis is to propose a new approach to "developmental evaluation" aimed at building a foundation for continuous improvement in large-scale school improvement networks, on the argument that doing so is essential to producing the intellectual capital needed to replicate effective practices and desired outcomes throughout these networks. We begin by developing a rationale for developmental evaluation, both to illuminate the need and to discuss its coordination with other forms of evaluation. We continue by proposing a logic of developmental evaluation to support analyzing networks as learning systems. We then use that logic to structure a framework for developmental evaluation to support evaluators, network executives, and other stakeholders in analyzing and strengthening the foundation for continuous improvement in a given network. Our analysis suggests that building a foundation for continuous improvement among a large number of networks is an educational reform agenda unto itself, one that must be supported and sustained if these networks are to succeed at the level expected under current accountability regimes.

¹University of Michigan, Ann Arbor, MI, USA
²George Washington University, DC, USA
³Education Trust–Midwest, Royal Oak, MI, USA

Corresponding Author:
Donald J. Peurach, School of Education, University of Michigan, 610 E. University, Ann Arbor, MI 48109, USA.
Email: dpeurach@umich.edu
Keywords
evaluation, developmental evaluation, impact evaluation, best practice, intellectual capital, educational reform, innovation, knowledge production, networks, organizational learning, replication, scale, sustainability

This is a conceptual analysis that addresses a practical problem: improving the production, use, and management of intellectual capital in the service of large-scale education reform.¹ By intellectual capital, we mean the practical, usable knowledge needed to coordinate and improve the performance of students, teachers, and school leaders in large numbers of schools. This intellectual capital is captured in formal resources (e.g., manuals, tools, digital media, and other artifacts); in individuals and relationships among them; and in relationships among schools and agencies, organizations, and constituents in their environments.

The analysis is grounded in a high leverage context: school improvement networks. These are new types of educational systems in which a central, “hub” organization collaborates with “outlet” schools to enact schoolwide improvement programs (Peurach & Glazer, 2012). Some of these networks operate outside of the K-12 governance structure: for example, networks operated by comprehensive school reform providers, charter management organizations, and education management organizations. Some operate within the K-12 governance structure: for example, “turnaround zones” in which newly constituted agencies centrally coordinate improvement efforts in large numbers of low performing schools.

Over the past 20 years, these networks have benefited from billions of dollars in public and philanthropic investment. Though some enlist functional schools seeking to transform existing capabilities, support is weighted heavily toward networks enlisting underperforming public schools and newly created charter schools, either to improve schools serving large populations of at-risk students or to create alternatives.

Given the accountability pressures under which they have emerged, the legitimacy and viability of school improvement networks are tightly tied to demonstrating levels of student performance on impact evaluations that many institutionalized educational systems have long struggled to obtain. Yet complex problems in schools, weak cause–effect knowledge on which to base programs, fledgling hub organizations, and turbulent environments interact to greatly reduce the prospects of quickly building large-scale networks of schools able to support students any better than institutionalized public schools (Berends, Bodilly, & Kirby, 2002; Center for Research on Education Outcomes,
Increasing recognition of the complexity and uncertainty of large-scale, systemic school improvement has been instrumental in motivating the Institute of Education Sciences (IES, 2013) to engage external evaluators, network executives, and other stakeholders in the continuous improvement of educational systems, including school improvement networks. This IES initiative is complemented by increasing use of design-based research to support the development of effective educational interventions (Anderson & Shattuck, 2012; Penuel, Fishman, Cheng, & Sabelli, 2011) and by emerging efforts to support continuous improvement in practice-focused educational networks (Bryk, Gomez, & Grunow, 2010).

New support for continuous improvement marks an advance in the national reform agenda, beyond a primary focus on impact to a complementary focus on producing, using, and managing the intellectual capital needed to demonstrate impact. However, in the case of school improvement networks, one problem is that there is little to suggest that hubs and schools have the foundation—the essential strategies and operational supports—to manage that which is to be continuously improved: intellectual capital. A second problem is that there is little to suggest that external evaluators, network executives, and other stakeholders are prepared to collaborate in improving intellectual capital in these novel and emerging systems.

The purpose of this analysis is to take up these two problems. We do so by proposing an approach to “developmental evaluation” aimed at establishing a foundation for continuous learning and improvement in school improvement networks. We begin by developing a rationale for developmental evaluation as a complement both to impact evaluation and to other improvement-focused evaluation strategies. We continue with a logic of developmental evaluation to support thinking and reasoning about networks as learning systems. We then use that logic to structure a framework for developmental evaluation to support evaluators, network executives, and other stakeholders both in critically analyzing networks as learning systems and in strengthening their foundation for continuous learning and improvement.

Our analysis provides novel perspective on school improvement networks as systems that produce, use, and manage the intellectual capital needed to improve education for many poor and at-risk students. Indeed, our analysis suggests that building a foundation for continuous improvement among a growing population of school improvement networks is an educational reform agenda unto itself, one that must be supported, sustained, and carefully managed if networks are to succeed at the level expected under current accountability regimes.
Rationale for Developmental Evaluation

We begin with a critical analysis of conventional means of evaluating school improvement networks. Our argument is that a disconnect between the conditions that would support success on widely required impact evaluations (on one hand) and the complex and uncertain conditions under which these networks actually operate (on the other) warrant commensurate support for developmental evaluation aimed at building a foundation for continuous learning and improvement within these networks.

Impact Evaluation, Its Goals, and Its Logic

What distinguishes school improvement networks from other large-scale reform strategies is that they take schools (rather than students) as the unit of intervention: not just their formal roles, structures, and technologies but also the teachers and leaders in schools, their individual capabilities and motivations, and their collective capabilities and culture.

The intervention, itself, is a comprehensive model for establishing and improving schoolwide operations. These models typically include complex organizational blueprints for structuring and restructuring schools, as well as designs for the practice (i.e., the day-to-day work) of leaders, teachers, and students. Enacting these designs for practice, in turn, requires intellectual capital: practical, useable knowledge as retained and elaborated in material, digital, and other resources; as manifest in individual teachers, leaders, external coaches, and their communities of practice; and as shared through practice-based learning opportunities. This intellectual capital is increasingly recognized as an essential resource for effecting coordinated improvements in leadership, instruction, and student achievement (Aladjem & Borman, 2006; Camburn, Rowan, & Taylor, 2003; Cohen & Ball, 2007; DeArmond, Gross, Bowen, Demeritt, & Lake, 2012; Rowan, Correnti, Miller, & Camburn, 2009a, 2009b).

These networks are emerging in policy contexts that hold them increasingly accountable for quickly establishing program impact. Rationale for doing so include assuring due diligence in the use of formidable public and private investment; stimulating competition for students and funding based on effectiveness; and recognizing the potential consequences of these networks for the lives of many students, teachers, and school leaders (for better or worse).

Impact evaluation typically has two goals (Raudenbush, 2007; Slavin & Fashola, 1998). The first is to identify a “treatment effect” evidenced by a positive, statistically significant difference in outcomes between students in
participating and nonparticipating schools, with more rigorous evaluations seeking to establish a causal relationship between the treatment (i.e., the schoolwide model) and outcomes. The second is to identify whether the treatment effect can be replicated beyond early adopters and in a broader pool of schools.

Replicable treatment effects are increasingly examined using a four-stage, “tiered evidence” sequence that culminates with impact evaluation. Each stage marks an increase in available funding, the number of participating schools, the standards of evidence, and, thus, the costs and sophistication of evaluation. Each stage also marks movement from formative to summative evaluation: that is, from evaluations that inform the incremental improvement of the schoolwide model to evaluations that determine its replicable effectiveness. 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 evaluation sequence along a 7- to 14-year timeline:

1. Evaluate a proposed program for its use of scientifically based research or other sources of “best practice” (1-2 years, preimplementation);
2. Implement in one or a small number of schools to establish “proof of concept,” with success evidenced via descriptive and other qualitative studies (1-3 years);
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 outcomes (2-4 years);
4. Further increase the installed base and use even more rigorous methods (e.g., quasi-experimental designs, randomized control trials, and meta-analyses) to further examine the magnitude and significance of effects (3-5 years).

This four-stage evaluation sequence is coupled closely with assumptions that the development of effective, replicable programs adheres to a sequential “RDDU” logic: research, development, dissemination, and utilization (Rowan, Camburn, & Barnes, 2004; see, also, Rogers, 1995). Basic and applied research feed development and small-scale pilots, from which follow rapid and widespread dissemination and effective use. This sequential, diffusion-centered logic model is highly institutionalized: for example, as evidenced by the use of this logic as the basis for the four-phase progression of the New American Schools initiative (Bodilly, 1996) the current goal
structure of the IES (2012b), and the three-stage “development/validation/ scale up” sequence within the federal Investing in Innovation (i3) program (U.S. Department of Education, 2010).

**Questionable Assumptions**

The RDDU logic is, itself, based on a set of assumptions about conditions that would enable the rapid development of effective, replicable programs in a 7- to 14-year window:

- Clear, shared understandings of the problems of (and goals for) schools.
- A knowledge infrastructure that includes a basic research enterprise providing robust, cause–effect knowledge; an applied research enterprise providing useable, research-based and research-validated components; and a professional education system that produces human resources capable of developing, supporting, and using these components.
- The possibility of hubs working with small numbers of schools in a short period of time to integrate previously tested and newly developed components into a generally effective schoolwide improvement model.
- The possibility of rapidly and faithfully transferring an established, multicomponent, schoolwide program to large numbers of schools that can quickly incorporate and use it to effect intended outcomes.

The problem, however, is that these conditions rarely (if ever) hold in practice, thus complicating efforts both to rapidly develop and scale up schoolwide models and to evaluate their impact. For example, the most relevant goals for schools are those in state accountability schemes (and not evaluation designs). Yet these schemes differ among states, are variably developed by content area and grade level, and are still evolving. Moreover, improvement goals are continuously redefined school-by-school, in response to the past performance of specific subgroups of students, specific content areas and grade levels, and even individual teachers and students. Consequently, the uses to which schools put a school improvement program are likely to differ in significant ways, with hubs as accountable for supporting differentiated use to meet school-specific goals as for maintaining program fidelity to meet evaluation goals.

The knowledge infrastructure in education is equally problematic. The basic research infrastructure has long been characterized as weak and
disconnected from practice (Bryk, 2009; Kaestle, 1993). As a locus of applied research and a possible source of validated program components, the “school improvement industry” is sprawling, turbulent, and dominated by conservative commercial publishers (Rowan, 2002), and it is just now coming under the oversight of emerging quasiregulatory agencies such as the What Works Clearinghouse and the Best Evidence Encyclopedia. And the professional education of teachers and school leaders has long been criticized as weak, lacking a coherent knowledge base, and uncoordinated with specific curricula, assessments, and other resources of practice (Levine, 2005, 2006; Sykes, Bird, & Kennedy, 2010).

Even if the knowledge infrastructure was robust, the use of small-scale pilots to devise a generally effective schoolwide improvement model is complicated by interdependencies in and among schools, the models, hub organizations, and the environments in which they operate (Cohen et al., 2014; Glazer & Peurach, 2013; Peurach, 2011). Each of the preceding consists of multiple and ever-evolving components, with difficult-to-discern relationships among them. Understanding, improving, and coordinating their interactions is difficult to accomplish in a small number of schools, and grounded in the particulars of pilot sites, their environments, and time. Indeed, networks often move from small-scale pilots to large-scale operations with programs that are promising-but-problematic, and under constant revision (Berends et al., 2002; Cohen et al., 2014; Glennan, Bodilly, Galegher, & Kerr, 2004; Marsh, Hamilton, & Gill, 2008; McDonald, Klein, & Riordan, 2009; Peurach, 2011).

Finally, rapid, effective, large-scale use is likely to be complicated not only by the scope, complexity, and uncertainty of schoolwide programs but, also, by problems and shortcomings in common strategies for diffusion and utilization. For example, strategies that emphasize such formal resources as codified routines and guidance have long been interpreted as a bureaucratic affront to local control and professional autonomy and, thus, resisted (Peurach, 2011; Rowan, 1990). At the same time, strategies that emphasize such social resources as mentoring, coaching, and communities of practice are limited by the geographic distance between hubs and schools; cultural and logistical obstacles to moving staff among classrooms and schools; small ratios of experts to novices; personnel transiency; and variability in local environments (Cohen et al., 2014; Peurach, 2011).

The Learning Imperative

Rather than enabling conditions supporting the RDDU sequence, longitudinal research on comprehensive school reform suggests that school improvement networks emerge under complex and uncertain conditions that challenge
operating in accord with the RDDU sequence (Berends, Bodilly, & Kirby, 2002; Bodilly, Glennan, Kerr, & Galegher, 2004; Cohen et al., 2014; Honig, 2006; Peurach, 2011). From this complexity and uncertainty follows a learning imperative: a need for hubs and schools to work less linearly and more circuitously, by collaborating over time to create, use, and refine practical knowledge supporting replicable effectiveness.6

Researchers describe this collaborative learning in terms of two interdependent, iterative learning processes: exploration and exploitation (Hatch, 2000; Peurach & Glazer, 2012). Exploration is a type of divergent learning that involves reconsidering premises and identifying new possibilities through search, experimentation, discovery, and invention.7 Exploitation is a type of convergent learning that involves leveraging established knowledge, selecting from among alternatives, and learning and refining through repeated use.

In that hubs and schools are learning through collaborative, iterative exploration and exploitation, the primary intervention in a school improvement network—the model for establishing and improving schoolwide operations—is not a fixed, objective, and effective “treatment” developed in advance of large-scale implementation. Rather, researchers have reconceptualized schoolwide improvement programs as subjective realities created through processes of co-construction and sensemaking among schools, districts, hubs, and other vested organizations in the context of large-scale implementation (Datnow, Hubbard, & Mehan, 2002; Datnow & Park, 2009). Effectiveness, in turn, depends both on leveraging established knowledge and on taking ownership and asserting agency in adapting and using knowledge in specific, local contexts (Coburn, 2003; McLaughlin & Mitra, 2001; Peurach & Glazer, 2012).

Consequently, knowledge of “best practice” does not exist in advance of scaling up, such that it can be readily incorporated and integrated into generally effective school improvement programs. Furthermore, knowledge of “best practice” does not remain constant over time, such that it will maintain currency despite ever-changing organizational and environmental contexts. Rather, increasingly better knowledge of practice emerges through the process of scaling up. This knowledge is intellectual capital that emerges (and is retained) among individuals and communities of practice within the network; that is captured and retained in codified resources, digital resources, and other artifacts; and that is adapted and refined over time through individual and collective use (Peurach & Glazer, 2012).

**Learning to Learn**

Although researchers have reported instances of this type of learning activity among a small number of school improvement networks, it is not safe to
assume that this is modal practice among the population of school improvement networks. Indeed, it is unlikely that all recognize the need for continuous learning and improvement, are equally adept, and work explicitly and proactively (rather than tacitly and reactively) to align their strategies, operational capabilities, and cultural norms to support the production, use, and management of intellectual capital.8

Consider hub organizations. These are often start-up enterprises with few demonstrated capabilities, founded and managed by educators, advocates, and others with little (if any) prior professional training or experience operating large-scale, knowledge-intensive organizations or networks. The development of such knowledge and capabilities among network executives is limited by weaknesses in available knowledge and professional learning opportunities to support their work.9 It is also limited by incentives and sanctions that often drive network executives to feign (if not actively pursue) operating in accordance with the RDDU logic.10 Indeed, for executives, upholding the myth of rationality is far safer than acknowledging complexity and uncertainty, and instrumental in maintaining legitimacy among funders and clients.

Consider, also, the chief collaborators of hub organizations: (a) existing schools with histories of underperformance and nonimprovement and (b) newly created schools with no history of past performance or improvement. These schools are likely to lack two capabilities essential to continuous learning and improvement via exploitation and exploration: absorptive capacity (i.e., the capability to leverage existing capabilities to recognize, value, incorporate, and use new practices and understandings) and dynamic capabilities (i.e., the capability to systematically generate and modify practices and understandings in pursuit of improved effectiveness, continued legitimacy, and sustainability).11 Moreover, efforts to develop these capabilities are complicated by labor market dynamics that have disproportionate numbers of weakly prepared teachers and school leaders working in underperforming schools, as well as by high rates of personnel transiency in underperforming and charter schools.12 The result is often the steady loss of newly created intellectual capital and the steady incorporation of weakly prepared teachers and school leaders.

Thus, even though the conditions under which they operate are likely to occasion the imperative to learn through interdependent exploration and exploitation, engaging that learning imperative is likely to require that school improvement networks learn to learn. That is, these networks must learn to develop and leverage the foundation—the essential strategies, operational infrastructure, and normative infrastructure—needed to create, use, retain, and manage intellectual capital through continuous learning and improvement.
That, in turn, begins with developing among network executives the understandings and capabilities needed to structure and manage school improvement networks as distributed, collaborative learning systems. Though not impossible, it is unlikely that many network executives will be able to independently develop knowledge and capabilities that, among innovating enterprises more broadly, have been found to be tacit, subconscious, and scarce. Rather, supporting network executives in learning to learn is likely to require the assistance and support of an external evaluator of some sort (e.g., a researcher, executive coach, and/or experienced mentor) sufficiently knowledgeable and skillful to provide guidance yet sufficiently humble to value and learn from the experiences of network executives.

The Case for Developmental Evaluation

Thus, school improvement networks operate in environments that link their legitimacy, funding, and sustainability to a progression of increasingly rigorous impact evaluations. However, conditions that would enable these networks to rapidly demonstrate replicable effectiveness at a large scale simply do not hold in practice. Instead, complex and uncertain conditions create a learning imperative: a need for hubs and schools to work collaboratively to produce, use, and manage intellectual capital supporting replicable effectiveness. Yet few networks are likely to have the foundation needed to support such learning, and few network executives are likely to know (or to independently learn) how to establish such a foundation. As such, many executives and networks will likely need support in learning to learn.

Developmental evaluation is an approach to evaluator/innovator collaboration with potential to address this need (Dozois, Langlois, & Blanchet-Cohen, 2010; Gamble, 2008; Patton, 2006, 2011, 2012). Consistent with the preceding analysis, developmental evaluation is grounded in assumptions that large-scale social innovations emerge and operate under conditions of complexity and uncertainty that challenge rational management and decision making and that require continuous learning and improvement. A chief aim of developmental evaluation, thus, is to support the development of large-scale social innovations through learning-centered, improvement-focused evaluation.

Central to developmental evaluation is the understanding that evaluation thinking, methods, and use need to be stitched deeply into the enterprise and integral to its management (in contrast to evaluation operating as a parallel process running alongside the enterprise). A key means for achieving this “stitching in” is for evaluators, themselves, to become integral members of the enterprise, working alongside social innovators and collaborating as
partners in their work (Dozois et al., 2010; Gamble, 2008; Patton, 2006, 2011, 2012). As a collaborative partner, the evaluator both contributes technical expertise and rigor to inquiry and analysis and, also, functions as a knowledgeable-and-critical friend who explicates the tacit, challenges assumptions, raises questions, and documents (and reexamines) decision-making processes in light of outcomes. Reciprocally, the evaluator leverages the experience to advance his or her understandings of the innovating enterprise, of social innovation more broadly, and of enacting the role of developmental evaluator.

An especially critical focus of developmental evaluation is the practice-based coaching and support of decision makers as they guide the enterprise, itself: that is, as they assess prevailing conditions; adapt and reconcile mission and strategy; align operations with mission and strategy; and build the understandings, commitment, and motivation of others (Dozois et al., 2010; Gamble, 2008). Because the complexity and uncertainty of prevailing conditions are likely to reduce the quality of available information, a primary role of the evaluator is to support decision makers both in interpreting partial and equivocal information and in communicating interpretations (and their meaning) to others. As such, key responsibilities of evaluators include framing and conceptualizing important issues and problems; identifying a parsimonious set of key indicators to guide rapid data collection; introducing frameworks and questions to guide interpretation and meaning-making; and building consensus among decision makers and others.15

Grounding principles of developmental evaluation in the preceding analysis of school improvement networks, a first-order matter of developmental evaluation early in the emergence of the network is for evaluators to support executives in learning to learn: that is, in learning to develop and leverage the foundation needed to produce, use, and manage intellectual capital through continuous learning and improvement.16 In doing so, the primary unit of analysis would not be the schoolwide program, and whether it works reliably and effectively. Rather, the primary unit of analysis would be the school improvement network and whether it is working as a learning system. Such analysis, in turn, would benefit from two key resources:

- A shared logic of developmental evaluation to support new ways of thinking and reasoning about networks as evolving through ongoing, iterative exploration and exploitation (and not through an RDDU sequence).
- A framework for developmental evaluation: that is, a parsimonious set of guiding questions to assess the network as a learning system; an interpretive framework to support analysis of information generated
using these indicators; and reflective questions to guide the assessment of possible implications.

With its focus on improving the network as a learning system, developmental evaluation would serve as a complement to other approaches to improvement-focused evaluation, by establishing a foundation that would allow these other approaches to be leveraged to greater effect. This includes design-based implementation research focused on the continuous improvement of programs and interventions (e.g., Anderson & Shattuck, 2012; Penuel et al., 2011). It also includes design-educational engineering-development focused on addressing problems of practice in network-based improvement communities (Bryk, 2009; Bryk et al. 2010; Mehta, Gomez, & Bryk, 2012).

By comparison, all three approaches to improvement-focused evaluation exist in a complex relationship with impact evaluation. Although improvement-focused evaluation has potential to increase prospects for success on impact evaluations, it also intentionally disrupts the many things that evaluators would hope to control in rigorous impact evaluations (none the least of which is the ostensible “treatment”). As such, any assessment of impact would be no more than a point estimate of a set of momentary conditions. For networks simultaneously engaged in both improvement-focused and impact evaluations, the relationship between them should be recognized as a tension to be managed and a key consideration in interpreting results.

A Logic for Developmental Evaluation

We continue by proposing a logic for developmental evaluation. The intent is to support external evaluators, network executives, and other stakeholders in thinking and reasoning more carefully about ways in which exploration and exploitation can interact to support the production, use, and management of intellectual capital, both at a large scale and in ways that ultimately have potential to support replicable effectiveness.

Specifically, we review and extend the evolutionary logic of replication advanced by Peurach and Glazer (2012). The logic draws on several knowledge-based traditions of organizational scholarship, including leading research on franchise-like organization replication in the commercial sector: an approach to large-scale organizational development that closely parallels that of school improvement networks. As applied to school improvement networks, the logic was initially used to structure an account of a leading comprehensive school reform program (Success for All) as a learning system and, then, refined through continued use and scholarship.
The logic is not intended as a “how to” prescription but, instead, as an ideal type: a heuristic for critically analyzing the foundation for continuous learning and improvement in specific networks, and for considering ways in which to build and strengthen that foundation. It is an alternative vision and a common platform on which external evaluators, network executives, and other vested parties can base the work of developmental evaluation, itself likely to evolve through use, reflection, and refinement.

**Review: The Evolutionary Logic of Replication**

As with school improvement networks, the evolutionary logic begins with a central, hub organization aiming to replicate a common organizational model across large numbers of outlets. The organizational model 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).

Recognizing the impossibility of creating precise organizational replicas in widely varying contexts, replication is considered successful when broadly equivalent outcomes are realized by similar means, with specific tolerances established within individual replication initiatives (Baden-Fuller & Winter, 2012). This approach has advantages in terms of speed, efficiency, and effectiveness under complex and uncertain conditions as described above: for example, weakness in knowledge and component technologies in broader environments; weakness in the knowledge and capabilities of outlet staff; and limits on the social management of knowledge through apprenticeship, mentoring, and communities of practice.

**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 physical characteristics, formal structures, or culture, simply because it is possible to replicate broad organizational forms without replicating organizational effectiveness (Winter & Szulanski, 2001). 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 toward intended outcomes than would be possible if outlets were working independently.

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. Reasons include
uncertainties, shortcomings, and flaws in available knowledge; inaccuracies and uncertainties in communicating complex practices and understandings; 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 within and among hubs and outlets.

**Foundations: Essential Knowledge Base and Core Learning Processes**

Given the preceding, the primary focus of the evolutionary logic is the production, use, and management of an essential knowledge base that supports the broad scope replication of capabilities. Consistent with research on school improvement networks, this essential knowledge base is produced, used, and managed through multiple iterations of two interdependent learning processes coenacted by hubs and outlets: exploitation and exploration (Winter & Szulanski, 2001; see also Bradach, 1998; March, 1996). This essential knowledge base is the core intellectual capital of the enterprise: a nonrivalrous resource that can be used repeatedly across many outlets without limiting its use in others.20

The essential knowledge base consists of three categories: knowledge of what, how, and where to replicate (Winter & 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).

**Emergence: A Template**

To establish proof of concept, development of the essential knowledge base begins with the construction of a “template”: an initial outlet that serves as a working example of the production or service capabilities to be replicated, often constructed in carefully selected sites and staffed with carefully selected people (Baden-Fuller & Winter, 2012; Winter, 2010; Winter & Szulanski, 2001). The template functions as a context for initial, exploratory learning in which hub and template staff engage in joint search, experimentation, discovery, and invention to devise means of realizing intended outcomes.
With successful exploration, the template becomes a repository of knowledge that the hub can study to develop provisional understandings of the capabilities 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 to be replicated across outlets: a plan describing intended activity in outlets, as well as a schema around which to develop and organize knowledge of practice. At a minimum, a design for practice describes essential roles, along with the “in principle” responsibilities of each role. A more developed design for practice can include qualifications for people occupying specific roles; descriptions and principles that detail the coordination among roles; and rubrics, goals, and standards for evaluating performance (both of individual roles and of the outlet as a whole).

**Essential Resource: Formalized Knowledge**

As the template matures, it becomes a context for the social, interpersonal management and reproduction of knowledge through apprenticeship, mentoring, and communities of practice. But, again, the use of social mechanisms to support large-scale organizational replication is limited by such issues as the broad scope of the knowledge to be replicated, geographic distances between the template and new outlets, logistical obstacles to moving staff between the template and new outlets, and small ratios of experienced to novice staff members.

As such, 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 tools, manuals, training materials, digital media, and other artifacts (Winter & Szulanski, 2001, 2002). The formalization of knowledge functions as a principal strategy for retaining, managing, and exploiting knowledge beyond the template and throughout the network.

Rather than as a “coercive” mechanism for exercising tight control over outlets, formalized knowledge is viewed as an “enabling” resource intended to support outlet staff in effectively performing coordinated work that would otherwise be beyond their immediate capabilities. Furthermore, although formal knowledge can be easily transferred or communicated to outlets, the assumption is that using this knowledge to recreate capabilities in outlets will require opportunities to learn about it and to practice using it.

Formalized knowledge falls into two categories. The first category 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). Routines are considered the primary mechanisms for supporting levels of coordinated activity that would otherwise be
dif
cult and costly to achieve (Nelson & Winter, 1982). 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 evaluate performance and outcomes. And they include “learning” routines that detail cycles of diagnosis, planning, implementation, and reflection.

The second category is codified guidance: professional and background knowledge essential to the understanding and enactment of specific roles and responsibilities, along with evaluation rubrics and decision trees that support analysis and decision making. Such guidance supports the intelligent (rather than rote) selection and enactment of routines, responsiveness to local circumstances, and the management of inevitable breakdowns and limitations in routines.

**Endemic Complication: Partial and Problematic Knowledge**

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

Consequently, hubs and outlets satisfice: that is, they commence replication with potentially rich (but partial-and-problematic) knowledge of key practices and understandings to be replicated in outlets, and (absent any experience replicating) with only speculative knowledge about where and how to replicate them. Consider the impossible alternative: 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 creating outlets and proceeding to large-scale replication. The aim is to recreate conventional capabilities for achieving common performance levels among outlets while
The developmental sequence begins with fidelity of implementation: exploiting knowledge by supporting outlets in learning to enact formalized routines as specified, with the goal of establishing conventional, coordinated, base-level capabilities and performance levels. Despite shortcomings and problems in the essential knowledge base, and despite the deferred benefits of addressing outlet-specific exigencies, fidelity provides multiple advantages. These include mitigating against weak initial capabilities in outlets; taking advantage of lessons learned and problems solved; creating opportunities to learn by doing (e.g., to enact practices, examine underlying principles, and examine the interdependence and coordination of activities); forestalling early problems (e.g., regression to past practice and 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 begin learning via exploration: that is, by assuming ownership and asserting agency in enacting the model to compensate for shortcomings, address problems, and respond to local needs and opportunities. Learning via 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. Capabilities for adaptive use are not assumed. Rather, the hub enables such activity using a collection of resources. These include open routines that support local decision making; assessment routines for evaluating performance and outcomes; “learning routines” that guide analysis, evaluation, and reflection; guidance that provides theories, principles, goals, standards, and other information to support and constrain local analysis, invention, and problem solving; and support for learning to use these resources.

The enactment of this developmental sequence also creates opportunities for the hub to engage in its own exploitation and exploration to refine and extend knowledge of where and how to replicate. For example, learning where to replicate involves using (and adapting) formal routines and guidance for identifying new outlets prepared for initial implementation, as well
as experienced outlets prepared to advance to adaptive use. Learning how to replicate involves using (and adapting) routines and guidance for use by coaches and trainers in supporting both base-level operations and adaptive use in outlets.

**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 exploitation: fidelity of implementation within and between outlets to establish conventional, base-level capabilities and performance levels. As they advance to exploration and 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, evaluates, and refines potential improvements; squares those with existing or new knowledge, resources, and requirements in broader environments; retains improvements both by incorporating them into an evolving template and by formalizing them as designs for practice, routines, and guidance; and works to purge ineffective practices. New practices and understandings are then fed back into existing outlets as incremental, “small-scope” improvements, and they are incorporated into a broader-yet knowledge base for use in creating 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 of exploitation and exploration result in an increasing (and increasingly refined) formal knowledge base detailing where, what, and how to replicate.

**Essential Mechanisms: Dynamic Capabilities**

Knowledge evolution is dependent on dynamic capabilities: learned routines through which hubs and outlets systematically generate and modify practices and understandings 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: systematic, disciplined, exploratory learning anchored in Deming-like “plan-do-check-act” continuous improvement cycles, with adaptations evaluated in light of outcomes and refined accordingly (Dosi et al., 2001). In hubs, dynamic capabilities are anchored in infrastructure and
capabilities for rapidly pooling and analyzing information and knowledge throughout the network; for evaluating the relationship between practices and understandings (on one hand) and intended outcomes (on the other); for experimentation and rapid prototyping; and for disseminating program improvements through the installed base of outlets by formalizing essential knowledge and supporting both faithful and adaptive use.

The more developed the dynamic capabilities in hubs and outlets, and the more they include formal and rigorous methods of internal evaluation and refinement (e.g., via design-based implementation research), the more likely that the knowledge evolution cycle will not only yield agreed on “best practices” but, instead, evidence-based practices linked empirically to relevant outcomes. Even so, extensive iterations will not yield omniscience. The essential knowledge base will always be partial and problematic, key knowledge will always remain undiscovered and/or tacit, and broader conditions are apt to change.

As such, knowledge evolution featuring iterative, interdependent exploitation and exploration functions as the essential capability of network-based organizational replication initiatives: a condition of “perpetual beta” enacted jointly by hubs and outlets over the life of the enterprise to support the production, use, and management of intellectual capital.

A Framework for Developmental Evaluation

We complete our analysis by proposing a framework for developmental evaluation. While establishing a rationale and logic for developmental evaluation is critical, building a foundation for continuous learning and improvement is, ultimately and essentially, practical work. As such, our aim is to build on our rationale and logic to provide actionable guidance to support researchers, network executives, and other stakeholders in critically analyzing the foundation for continuous learning and improvement in a given school improvement network. The framework consists of three components: five guiding questions to structure data collection, a four-category interpretative framework to support the analysis of data generated using these questions, and three questions to structure collective reflection.

Guiding Questions

We begin by adapting guiding questions first proposed by Peurach and Glazer (2012) and Peurach, Glazer, and Lenhoff (2012). These five questions are intended to structure the rapid collection of a parsimonious-yet-powerful body of evidence (through interviews, document analysis, and participant
observation) about characteristics of the network suggested by the evolutionary logic as foundational to continuous learning and improvement. The first question examines the alignment between the network’s strategy for managing intellectual capital and the complex and uncertain conditions under which networks are likely to operate. The remaining four questions examine the alignment between that strategy and internal operations.

1. **Does the enterprise have an explicit strategy for managing intellectual capital that attends to both exploitation and exploration?** Attention to exploitation is evidenced by goals, norms, and language that emphasize the faithful implementation of evidence-based (or otherwise-established) practices. Attention to exploration is evidenced by goals, norms, and language that emphasize local experimentation, invention, and adaptation.

2. **Does the enterprise have a formal design for practice?** Akin to an explicit program theory or logic model, such a design would be evidenced by formal descriptions of essential roles; qualifications for essential roles; principles detailing responsibilities and coordination among roles; and standards and rubrics for assessing the enactment of those roles. It would be further evidenced by a functional template from which the design was drawn (and in which it can be observed and studied in operation).

3. **Does the enterprise feature formal, codified resources for recreating base-level practices and understandings in schools?** These resources are evidenced by formal routines and guidance for recruiting, selecting, and enlisting schools in which conditions exist (or can be created) to support base-level operations; by formal routines and guidance for use by schools to establish consistent, base-level practices and understandings; by formal routines and guidance for use by trainers and coaches to support schools in establishing base-level practices and understandings; and by language and frameworks to guide the interpretation of these resources as enabling (and not coercive).

4. **Does the enterprise feature formal, codified resources for recreating practices and understandings 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 mastered base-level operations (and, thus, are prepared to progress to adaptive use); by formal routines and guidance for use by school staff to support design, evaluation, problem solving, decision making, and other discretionary activity; by formal routines and guidance for use by
trainers and coaches to support discretionary activity in outlets; and by language and frameworks that encourage divergence while also minding conventions.

5. *Does the hub organization have the infrastructure and capabilities to support evolutionary learning?* Such infrastructure and capabilities would be evidence by the above-described supports for adaptive use (as a source of within-network variation in practices and understandings) and for base-level capabilities (to feed program improvements back through the network). They would be further evidenced by a communication infrastructure supporting the bilateral exchange of knowledge and information among hubs and schools; opportunities, resources, and capabilities in the hub for analyzing school performance and outcomes (as via design-based implementation research); and opportunities, resources, and capabilities for formalization, rapid prototyping, and small-scale evaluation.

**Interpretive Framework**

Our conjecture is that few school improvement networks will be fully attentive to exploitation and exploration in ways suggested by the evolutionary logic. After all, these networks operate amid a legacy of past educational reform movements that were vigilant on matters of structural compliance but surprisingly inattentive to the learning required to effect complementary changes in practice. Furthermore, they operate in reform environments that have long understood exploitation and exploration as mutually exclusive and ideologically steeped alternatives, with faithful implementation of external guidance understood both as coercive and as fundamentally at odds with local and professional autonomy, invention, and design. Finally, they operate amid institutionalized understandings of innovation as an RDDU sequence, absent understandings of either the possibility or legitimacy of integrating exploitation and exploration to support continuous learning and improvement.

As such, we continue by proposing an interpretive framework to differentiate among networks by their foundations (i.e., their strategies and supports) for continuous learning and improvement. Using evidence generated with our guiding questions, the framework identifies networks as structured and operating consistent with one of four primary types: a shell enterprise, a diffusion enterprise, an incubation enterprise, or an evolutionary enterprise. The framework also identifies vulnerabilities among these types for the production, use, and management of intellectual capital, as well as implications for implementation and outcomes.
Shell enterprise. A shell enterprise is one in which the hub seeks to replicate distinguishing organizational characteristics across schools (e.g., roles, structures, tools, and/or culture) absent efforts to recreate essential capabilities. As such, a shell enterprise will be evidenced by detailed organizational blueprints, including (potentially) a formal design for practice. However, a shell enterprise will show little or no evidence of an explicit strategy for managing intellectual capital, as well as little or no evidence of formal supports for either base-level operations (i.e., exploitation) or adaptive use (i.e., exploration).30

Diffusion enterprise. Consistent with the institutionalized RDDU logic, a diffusion enterprise is one in which the hub places a primary emphasis on codifying proven and/or established practices to be enacted with fidelity in schools. A diffusion enterprise is evidenced by a strategy for managing intellectual capital that focuses primarily on exploiting available knowledge of “what works.” It is further evidenced by (a) extensive, formal routines and guidance for creating consistent, base-level operations (i.e., exploitation) and (b) comparatively weak attention to supporting experimentation and adaptation (i.e., exploration).

Incubation enterprise. Mindful of local control and professional autonomy, an incubation enterprise is one in which the hub places a primary emphasis on structuring parameters, processes, and resources to support school-level design, implementation, and problem solving. An incubation enterprise is evidenced by a strategy for managing intellectual capital that focuses primarily on distributed, exploratory learning through which schools operationalize hub-formalized designs for (and principles of) practice. It is further evidenced by (a) extensive, formal routines and guidance supporting adaptive, locally responsive use (i.e., exploration) and (b) comparatively weak emphasis on (and support for) the faithful implementation of specific methods of production or service delivery (i.e., exploitation).

Evolutionary enterprise. An evolutionary enterprise is one in which hubs and schools engage in collaborative learning that yields a formal knowledge base detailing where, what, and how to replicate. An evolutionary enterprise is evidenced by a strategy for managing intellectual capital that emphasizes both exploitation and exploration, as well as by formal routines and guidance supporting both base-level operations (exploitation) and adaptive use (exploration).

Vulnerabilities and implications. From the perspective of continuous learning and improvement, shell, diffusion, and incubation enterprises lack mechanisms that would support the development of a formal knowledge base that could be used to recreate and refine capabilities for effective practice in large
numbers of schools. For example, shell enterprises lack the foundation for exploration (to establish base-level operations), exploitation (to support adaptation and problem solving), and iterations between the two (to continuously learn and improve). Furthermore, while diffusion enterprises establish mechanisms for exploiting established knowledge, they lack mechanisms to support exploration, adaptation, problem solving, and feedback that would introduce variation and new knowledge. Finally, while incubation enterprises support exploration, adaptation, and problem solving, they also lack mechanisms for culling, testing, and exploiting new knowledge as it emerges, and they run the risk of new knowledge being so context-specific as to have little value beyond individual schools.

Weaknesses in the resulting knowledge base, in turn, create risks for implementation and outcomes for networks operating under the uncertain and complex conditions described above; with schools likely weak in initial capabilities; with increasing scale likely straining social mechanisms for managing intellectual capital; and with impact evaluation likely on the horizon. For example, shell enterprises risk classic loose coupling, with weak linkages between the formal and behavioral structure of schools and, thus, a high risk of regression to past practice. Furthermore, diffusion enterprises risk schools faithfully enacting practices as specified, neglecting to address local needs, and thus capping program impact below desired levels. Finally, incubation enterprises risk both variation in the “treatment” and regression to past practice, thus complicating efforts to identify a “treatment effect.”

Failure isn’t inevitable. It is conceivable that schools could compensate for network-level weaknesses with strengths of their own: for example, prior knowledge and capabilities; the ability to incorporate and use other knowledge; and the ability to learn from experience. Yet, beyond the possibility of a small number of positive outliers, our earlier analysis suggests that chronically underperforming and newly created schools are likely to lack such capabilities.

The risk, then is of a Matthew effect: an asymmetric, bimodal distribution in implementation and outcomes determined by the initial capabilities of new schools, with a small number of initially capable schools succeeding and many others struggling. With that, the network’s strategy for managing intellectual capital becomes a chief source of precisely the type of variation that complicates establishing replicable effectiveness on impact evaluations.

**Structured Reflection**

Again, our conjecture is the networks are more likely to be identified as shell, diffusion, and incubation enterprises than as evolutionary enterprises. From the perspective of developmental evaluation, the task for external evaluators
is to use the preceding evidence and interpretations to motivate and guide critical reflection among network executives and stakeholders, with the goal of understanding (and, possibly, improving) the network’s foundation for continuous learning and improvement, given the likelihood of impact evaluation.

This is not a straightforward task, and would surely benefit from guidance all its own. It would require that external evaluators, executives, and stakeholders understand the rationale, logic, and framework for developmental evaluation. Furthermore, it would require collectively reviewing evidence of the network’s current strategies and supports, reaching consensus on interpretations, and squaring hypothesized vulnerabilities and risks with evidence of implementation and outcomes. Finally, it would require developing shared understandings of what are likely to be multiple, mutually reinforcing conditions that have the network pursuing a shell, diffusion, or incubation strategy: for example, the understandings, ideologies, beliefs, and identities of network executives, as well as those of the people and schools that have joined the network; the expectations and constraints of funders and other stakeholders; institutionalized understandings and ideologies of education and its reform; and limitations in resources.

Assuming the development of such understandings, we conclude by providing guidance to structure collaborative reflection about possible next steps. Specifically, we consider three options: staying the course, transforming the network, and learning to learn. The catch is this: While each is a legitimate possibility, each also presents its own problems.

Option 1: Staying the course. For networks vested in shell, diffusion, or incubation strategies, one possible response is to stay the course: that is, to hold tight to the existing strategy while, at the same time, trying to reduce the complexity and uncertainty under which the network operates. This approach might include working in comparatively developed domains of research and professional preparation (e.g., early reading); working with more established, able, and stable schools; working at a scale that permits the social management of knowledge; and avoiding circumstances that call for impact evaluation.

However, it is not clear that any domain of educational activity is sufficiently developed as to obviate the need for both exploitation and exploration. Furthermore, efforts to establish favorable conditions (e.g., through detailed program adoption processes, district-level collaboration, and broader lobbying efforts) are often ineffective (Datnow, 2000), and policy-level efforts to establish coherent environments often do more to effect turbulence than to effect enabling conditions (Glazer & Peurach, 2013; Trujillo,
Finally, the stronger the schools, the smaller scale, and less the concern with replicable effectiveness, the greater the risk that the network will define itself out of the policy-supported and philanthropic-supported agenda for educational reform (and, with that, struggle to secure resources and schools).

**Option 2: Transforming the network.** A second option is for executives to commence a rapid and radical transformation of the network to reorganize as an evolutionary enterprise. Diffusion enterprises would place renewed emphasis on exploration, incubation enterprises would place renewed emphasis on exploitation, shell enterprises would attend to both, and all would place renewed emphasis on establishing centralized dynamic capabilities. This would have the advantage of aligning the network’s foundation for continuous learning and improvement with the complexity and uncertainty under which it is likely to operate, thus (in principle) increasing the potential of quickly producing, using, and managing the intellectual capital needed to demonstrate replicable effectiveness.

Even so, rapid transformation is risky. Again, the evolutionary logic is an ideal type to support analysis, and not a set of “how to” prescriptions. Furthermore, though research provides a number of accounts of networkwide learning via exploration and exploitation, it provides only one detailed account of a school improvement network operating in full accord with the evolutionary logic, the result of three decades of self-guided learning and improvement (Peurach & Glazer, 2012). Finally, research neither provides accounts of a school improvement network intentionally reorganizing as an evolutionary enterprise nor any knowledge of how to do so.

Rapidly reorganizing as an evolutionary enterprise would likely be exceedingly difficult, in that it would require fundamentally deconstructing and reconstructing the entire system of interdependent, mutually reinforcing conditions that support operating as a shell, diffusion, or incubation enterprise. For example, it would require that network executives have the confidence and humility to reconstruct their understandings, ideologies, and identities, at the same time they support hub and school staff in doing the same. Furthermore, it would require expanding the operating capabilities of the hub, recruiting staff members and collaborators with new knowledge and capabilities, and reducing other activities and reallocating resources. Finally, it would require that funders, policy makers, and other stakeholders not only reconstruct their own understandings but also create conditions that would support reorganizing as an evolutionary enterprise: for example, by legitimizing the learning imperative; building political support; and providing necessary time and resources.
Indeed, the uncertainty and challenges of rapidly reorganizing as an evolutionary enterprise could be so daunting that network executives decide, instead, to incur the vulnerabilities and risks of continuing to operate as a shell, diffusion, or incubation enterprise.

**Option 3: Learning to learn.** A third option is to learn to learn. This option would have executives, stakeholders, and evaluators initiating cycles of exploration and exploitation more limited in scope than networkwide transformation, with two goals: strengthening the knowledge needed to demonstrate replicable effectiveness and, importantly, building the knowledge needed to manage movement toward an evolutionary enterprise.

This approach would begin by reviewing the network’s design for practice to identify exceptional instances in which particular roles are operating in accord with the evolutionary logic, with a formal knowledge base accumulating to support both conventional, base-level operations and adaptive, locally responsive use. These exceptional instances would function as existence proofs of the possibility of pursuing an evolutionary strategy within the network, and of the conditions that support doing so.

Building on evidence of possibility, this approach would continue with again examining the design for practice to identify a small set of “linchpin roles” particularly instrumental to replicating effectiveness at scale: for example, coaching or leadership roles with responsibilities for supporting the performance of others. Attention would then focus on those linchpin roles for which role incumbents were likely to have weak initial capabilities, and for which increasing the scale of operations would likely tax social mechanisms for managing intellectual capital (e.g., apprenticeship, mentoring, and communities of practice).

The hub would then explore the possibility of supporting these roles in accordance with the evolutionary logic: that is, by codifying routines and guidance to support base-level enactment of key responsibilities; by codifying routines and guidance to support and bound discretionary enactment; by introducing these routines and guidance in developmentally appropriate ways, progressing from fidelity to adaption; and, finally, by monitoring implementation, evaluating and codifying new knowledge as it emerges, and diffusing it throughout the network. Knowledge gained by executives through reforming individual roles in this way could then be exploited to reform additional roles over time.

With that, learning to learn would, itself, be an evolutionary progression toward operating as an evolutionary enterprise, one that would heed argued advantages in improving the network’s foundation for continuous learning and improvement while, at the same time, managing the uncertainty and
challenges in doing so. The risk, however, is that this would be a slow process in policy and funding environments only now beginning to recognize the need for such learning; in which the myth of rationality has long held sway despite overwhelming evidence of complexity and uncertainty; and in which the continued viability of networks is linked tightly to rapidly demonstrating replicable effectiveness on rigorous impact evaluations.

**Conclusion: Reforming Practice and Practicing Reform**

Anchored in contemporary concern with the continuous improvement of school improvement networks, our analysis contributes to long-running traditions of policy, research, and reform focused on the role of knowledge and knowledge production in educational improvement. It also provides critical perspective on contemporary efforts to leverage research and evaluation in the service of large-scale education reform, including the Obama administration’s tiered evidence sequence for social innovation and its heavy focus on impact evaluation.

Our specific focus is reforming the practice—the day-to-day work—of students, teachers, and school leaders. Our analytic foil is the research–development–diffusion–utilization sequence: a highly institutionalized set of understandings about creating and using knowledge, and the paradigm that currently structures the leading funding opportunities and evaluation demands for large-scale educational improvement initiatives.

Recognizing both the imperative to evaluate school improvement networks and the challenges of doing so, we began with a rationale establishing the need for pursuing developmental evaluation as a complement both to impact evaluation and to other approaches to improvement-focused evaluation. We continued by proposing an evolutionary logic to support thinking and reasoning about developmental evaluation as focused on the production, use, and management of intellectual capital under conditions of complexity and uncertainty. We concluded by proposing a parsimonious framework to structure inquiry, analysis, and reflection about the current state of a school improvement network and about strengthening its foundation for continuous learning and improvement.

Our general argument is that the prospects for continuous learning and improvement increase with a positive coalignment among three things: the conditions under which a school improvement network operates; its strategies for producing, using, and managing intellectual capital; and its operational supports for collaborative learning among the hub and schools. For most school improvement networks, achieving such positive coalignment will require learning to
learn, with the conditions under which they are likely to operate driving networks to develop strategies and operational supports that combine exploitation and exploration in the service of two goals: building a formal knowledge base with potential to support replicable effectiveness; and building knowledge among executives, stakeholders, and evaluators to manage such work.

With that, our analysis suggests that reforming practice will require practicing reform: executives, stakeholders, and external evaluators learning about (and from) the work of producing, using, and managing intellectual capital in the context of large-scale educational networks. Doing so will require that these unlikely collaborators critically examine the fundamental conditions under which networks operate; their strategies for producing, using, and managing intellectual capital; and the assumptions, ideologies, and identities on which those strategies are based. It will also require that they make reasoned judgments about how best to move forward under conditions of complexity and uncertainty, and that they continuously reflect on their rationale and logic for moving forward in light of the experiences that follow.

One next step, then, is to begin experimenting with exactly that: actually using the proposed logic and framework to experiment with supporting a small number of school improvement networks open to analyzing and improving themselves as learning systems. Another next step would be to develop (and similarly experiment with) alternative logics and frameworks for improving networks as learning systems. Doing so would provide opportunity to create additional resources and guidance to support evaluators, network executives, and other stakeholders in collecting, analyzing, and reflecting on evidence about networks as a learning system. Furthermore, it would provide opportunity to refine the proposed logics and frameworks through use. Finally, it would provide opportunity to examine the feasibility, enabling conditions, and payoff for developmental evaluation, especially if carefully coordinated with complementary analyses of implementation and outcomes.

Another next step is to use the analysis developed here as a stepping stone into understanding continuous learning and improvement in other approaches to large-scale education reform. This includes approaches that use designs for improvement less comprehensive than the schoolwide models used in school improvement networks. It also includes approaches that leverage network forms other than the type of hub-outlet topology described here: for example, self-organizing networks that emphasize reciprocal relationships among schools absent a strong, coordinating hub organization.

As argued at the outset, we are edging toward a much fuller appreciation of the challenges of large-scale school improvement, and of the different forms of knowledge and knowledge production that support it. This fuller appreciation is captured clearly in IES moving beyond an essential focus on
program impact to an equally essential focus on the continuous improvement of educational systems.

Our analysis affirms the wisdom of these efforts and predicts challenges likely to follow. Demonstrating replicable effectiveness on rigorous impact evaluations depends on the type of continuous improvement initiatives that IES is just now beginning to legitimate and fund. Yet the success of these initiatives likely depends on foundational organization-building in novel educational systems; the success of those efforts depends on the practice-based professional development of novel types of educational leaders; and the success of those efforts depends on devising new types of frameworks and methods to support such learning, as well as new types of external evaluators to occasion and lead such learning. Furthermore, absent careful coordination, success along the preceding dimensions would likely have school improvement networks working at cross purposes, in that they would be learning and improving in ways that could well undermine rigorous, complex, controlled impact evaluations.

Thus, some might view new support for the continuous learning and improvement of educational systems as another plank in the RDDU/tiered evidence platform. Instead, we view this support fundamentally differently: as new, complex, and uncertain territory, the mapping of which will require that unlikely collaborators learn to think, reason, and coordinate in entirely new ways, both about the production, use, and management of intellectual capital and about efforts to evaluate impact.

The former perspective has continuous improvement research as part of the answer. The latter perspective has continuous improvement research as yet another piece of the puzzle, and a reform agenda all its own.

Acknowledgment

The lead author gratefully acknowledges the opportunity to participate in the i3 Learning Community cosponsored by the W. T. Grant Foundation and the Spencer Foundation, as well as the opportunity to discuss this work in the Seminar on the Evolution of Organizations and Industries at the Wharton School. Both experiences served as important contexts in which to further develop and advance this work.

Authors’ Note

An earlier draft of this article was presented at the 2012 Conference of the National Center on Scaling Up Effective Schools, Nashville, TN (June 10-12, 2012).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors gratefully acknowledge funding received from the Spencer Foundation (Reference No. 201300078), the School of Education at the University of Michigan, and the Education Policy Center at Michigan State University.

Notes

1. Our conceptualization of intellectual capital draws from Bontis (2002) and Stewart (1997). Bontis considers intellectual capital as the stock of knowledge of an organization and as the product of organizational learning over time. Furthermore, he describes intellectual capital as consisting of three subdomains: structural capital (i.e., the organizational routines of an organization), human capital (i.e., the tacit knowledge embedded in individuals and their interactions), and relational capital (i.e., the knowledge embedded in relationships established in outside environments). Stewart, in turn, argues that the value of intellectual capital lies in its practical use in creating instrumental outcomes (in his case, wealth; in our case, improved performance in schools).

2. For example, school improvement networks have emerged concurrent with the rise of standards-based reform in states; the establishment of the Institute of Education Sciences (IES; 2012a), and its mission of identifying “what works, what doesn’t, and why”; funding streams that require rigorous evaluation of outcomes; the coordination of evidence standards among federal agencies; and the rise of such quasiregulatory agencies as the What Works Clearinghouse and the Best Evidence Encyclopedia. Such efforts in education parallel wider efforts to establish the impact of other social programs, both in the United States and abroad (Campbell Collaboration, 2013; Granger, 2011; Khandker, Koolwal, & Samad, 2010).

3. This tiered evidence sequence is one component of the Obama administration’s efforts to advance the use of evidence in support of social policy and innovation—efforts described as by Haskins and Baron (2011) as the most expansive in the history of the U.S. government. For an overview of the Obama administration’s strategy for using evidence and innovation to improve government performance, see Burwell, Munoz, Holdren, and Krueger (2013) and IES/National Science Foundation (2013).

4. Time estimates are derived from IES (2012b).

5. Rather than being unique to school improvement networks, weaknesses in such enabling conditions are characteristic of innovation in general (Van de Ven, Polley, Garud, & Venkataraman, 1999) and of social innovation in particular (Preskill & Beer, 2012). Multiple efforts are underway to strengthen the impact evaluation infrastructure in education in ways that would support evaluating the impact of complex educational interventions amid weaknesses in enabling conditions: for example, the establishment of the IES in 2002; the establishment
of Society for Research on Educational Effectiveness (the chief professional organization focused on understanding cause–effect relationships in educational programs and interventions) in 2005; the development of methods such as regression discontinuity design (Schochet, 2008); efforts to conceptualize sources of variation in program effects (M. J. Weiss, Bloom, & Brock, 2013); and multiple IES-sponsored initiatives to develop capabilities of researchers to conduct randomized control trials, impact evaluations, and causal analyses. Even so, within education, issues related to the potential and problems of impact evaluation have been (and continue to be) hotly debated among proponents and critics (e.g., Foray, Murnane, & Nelson, 2007; Mosteller & Boruch, 2002; Schneider & McDonald, 2007). Moreover, 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 et al., 2010).

6. The learning imperative is not unique to school improvement networks. Rather, comparative research on the innovation process directly refutes the RDDU (research, development, dissemination, and utilization) sequence and, instead, argues that the innovation process is better understood as cycles of exploration/divergence and exploitation/convergence (Van de Ven, Polley, Garud, & Venkataraman, 1999). Such learning is argued to be the essential capability for innovating organizations operating in turbulent, uncertain, and chaotic environments (Lewin, 1999; Waldrop, 1992). This learning imperative is also captured in leading characterizations of “design thinking” (T. Brown, 2009), “perpetual beta” in the development of technical innovations (Musser & O’Reilly, 2006), and the development and refinement of “better” (rather than “best”) medical practice (Berwick, 2008; Gawande, 2002, 2007, 2009).

7. Exploration and exploitation have roots in March (1996). Furthermore, “exploitation” parallels “double loop learning” as developed by Argyris and Schön (1978), where “exploitation” parallels “single loop learning.”

8. In organizational studies, capabilities for continuous learning and improvement are recognized as unevenly distributed among organizations, and as a source of competitive advantage for innovating organizations operating in complex and turbulent environments (Choo & Bontis, 2002; Dosi, Nelson, & Winter, 2001).

9. For example, scholars describe a “paucity of research” on the work of hub organizations (Datnow, Hubbard, & Mehan, 2002, p. 90; see, also, Allen & Peurach, 2013; Peurach & Gumus, 2011). Furthermore, professional learning opportunities are limited to a small number of executive education programs (e.g., Harvard University’s Doctor of Education Leadership and the Broad Residency) and to small communities of practice coordinated by leading philanthropists (e.g., the i3 Learning Community, which is sponsored by the Spencer Foundation and the W. T. Grant Foundation; the Deeper Learning Initiative, which is sponsored by the William and Flora Hewlett Foundation).

10. For example, executives in charter management organizations (CMOs) report a “tyranny of business plans” resulting from the need to project rationality to secure
funding, despite encountering complexity and uncertainty that require flexibility and adaptability in their work (National Charter School Research Project, 2007). Again, such pressure is not unique to school improvement networks but, instead, characteristic of social innovation enterprises (Preskill & Beer, 2012).


12. For example, see Almy and Theokas (2010) and Ingersoll (2001) for analyses of occupational demographics. See Ronfeldt, Loeb, and Wyckoff (2013) for a review and analysis of teacher transiency (and its negative consequences for students and for entire schools). See Stuit and Smith (2009) and Gross and DeArmond (2010) for comparative research on teacher turnover in charter schools and conventional public schools.

13. For example, regarding the challenges of developing executive capacity to manage learning-focused innovation processes, see Van de Ven et al. (1999). Regarding challenges in strategically managing intellectual capital, organizational knowledge, and organizational learning, see Choo and Bontis (2002). Finally, regarding the emergence of (and challenges in) the role of “chief knowledge officer” as a formal executive role responsible for such work, see Earl and Scott (1999).

14. Developmental evaluation is an emerging approach, and not an established tradition or method. Its principles, central tenets, and unique contributions are developed most fully in Patton (2011), drawing on prior work on utilization-focused evaluation (Patton, 2008) and social innovation (Westley, Zimmerman, & Patton, 2007). Two practical guides for developmental evaluation have emerged out of early collaborations with the McConnell Family Foundation (Dozois, Langlois, & Blanchet-Cohen, 2010; Gamble, 2008). The notion of developmental evaluation has been quickly and widely embraced for supporting innovation. For example, a simple Google search using “developmental evaluation” and “innovation” yielded hundreds of sources. Even so, a search on “developmental evaluation” using Google Scholar yielded very little research on (or using) developmental evaluation in peer-reviewed journals (by our count, less than 10 in the first hundred sources identified at the time of this writing). Among both the peer-reviewed and non-peer-reviewed reports that we identified, most function as “proof of concept” affirming the principles and tenets of developmental evaluation, and most report positive value to program developers. These include one study focused on supporting a new approach for teaching classroom assessment methods to aspiring teachers (Lam, 2011), one focused on supporting learning in a network context (Ramstad, 2009), and one focused on supporting the development of interorganizational networks (Sydow, 2004). Our conjecture is that the small number of peer-reviewed studies arises, in part, from developmental evaluation being focused primarily on internal use within organizations (and not on testing and advancing general knowledge, at least beyond the principles and practices of developmental evaluation, itself). Furthermore, our experience is that quickly identifying peer-reviewed studies is complicated by the fact that
“developmental evaluation” is an established diagnostic regime in psychology, and the subject of extensive research.

15. As argued both by Gamble (2008) and Patton (2011), central to this interpretation is research on executive decision making in complex environments with high costs (and little potential) for accurate information and knowledge (Sutcliffe & Weber, 2003). As summarized by Sutcliffe and Weber, “Our findings suggest that perceptual accuracy at the very top executive levels is actually a source of competitive disadvantage for most firms. The task of leaders is to manage ambiguity and to mobilize action, not to store highly accurate knowledge about their environment. The more effective way to improve the performance of a company is to invest in how leaders shape their interpretive outlooks” (Sutcliffe & Weber, 2003, p. 82).

16. Working from this perspective, the meaning of “development” in “developmental evaluation” begins to shift. Specifically, it shifts away from a more behavioral connotation: the sort of active program fashioning connotated by research—development—diffusion—utilization and as captured by the “development” stage of the tiered evidence sequence. And it shifts toward a more cognitive connotation: establishing the resources and capabilities needed to support collective thinking, reasoning, and understanding. As argued by Patton (2011), this focus on continuous learning amid complex and uncertain conditions (vs. active program fashioning and related problem solving) is what distinguishes developmental evaluation from other improvement-focused evaluation strategies, including design-based research and action research.

17. As noted by Peurach and Glazer (2012), the evolutionary logic is drawn primarily from theory and research by Sidney Winter, Gabriel Szulanski, and colleagues focused on the replication of knowledge within and between organizations: for example, Baden-Fuller and Winter (2012); Szulanski and Winter (2002); Szulanski, Winter, Cappetta, and Van den Bulte (2002); Winter (2003, 2010, 2012); Winter and Szulanski (2001, 2002); and Zollo and Winter (2002). Roots of this work lie in the work of Nelson and 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, 1996); innovation development (Van de Ven et al., 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; J. S. Brown & Duguid, 1998; Eisenhardt & Martin, 2000; Grant, 1996; Wernerfelt, 1995); alternative conceptions of centralized control (Adler & Borys, 1996); franchised organizational forms (Bradach, 1998); and nonprofit replication (Bradach, 2003).

18. The evolutionary logic as represented here incorporates three important adaptations over the logic as originally represented in Peurach and Glazer (2012). The first adaptation is our naming of the logic. Where we initially represented it as a “knowledge-based” logic, we came to recognize that RDDU sequence is, itself, a knowledge-based logic, such that the “knowledge-based” qualifier did
not discriminate between the two logics on which we focus. Hence, we renamed this as an *evolutionary* logic, recognizing its deep roots in evolutionary economics. The second adaptation is our elaboration of conditions under which an evolutionary strategy has advantages in speed, effectiveness, and efficiency over a more straightforward “in principle” articulation of capabilities to be developed and coordinated in new outlets. The third adaptation is our incorporation of the notion of a “design for practice” as an essential component of the evolutionary logic, which becomes central to the interpretative framework developed later in this article. These adaptations were motivated by personal communications with Sidney Winter and Charles Baden-Fuller about their work on “principles” and “templates” as strategies for organizational replication; by feedback on our analysis from participants in the Seminar on the Evolution of Organizations and Industries at the Wharton School; and by efforts actually practicing the work of developmental evaluation in collaboration with a large-scale effort to support the implementation of Response to Intervention and Positive Behavioral Interventions and Supports in Michigan.

19. “Outlet” is the general term describing the organizations that are to be replicated. In school improvement networks, the outlets are schools.

20. Winter and Szulanski (2001) describe this knowledge base as the “Arrow core” in recognition of Kenneth Arrow’s (1962) exposition of information economics: in particular, his analysis of information as a nonrivalrous good, the fundamental assumption on which the evolutionary logic of replication rests. To say that this knowledge base is nonrivalrous is not to say that it is (or should be) available to other, possibly competing enterprises. An enterprise may well take measures to protect the use of this knowledge base through copyrights, trademarks, patents, noncompete agreements, and other means of protecting intellectual property rights. Indeed, whether the knowledge base produced by school improvement networks (through public funding, in the service of public education) is a public or private good is a formidable policy issue that arises from this analysis, especially because current policy contexts do more to structure competition among these networks (thus, the protection of intellectual capital, despite being produced with public funding for the public good) than collaboration (thus, the sharing of intellectual capital as a public good in the public domain).

21. The work of Winter, Szulanski, and colleagues generally places more emphasis on routines than on guidance. However, the importance of professional and background knowledge as a complement to routines becomes salient in Baden-Fuller and Winter (2012).

22. Knowledge thus formalized functions as a sort of “immutable mobile” (Latour, 1988) that can be used, studied, and manipulated by others.


24. As noted by Peurach and Glazer (2012), Szulanski et al. (2002) cast this as a four-phase process. As forms of exploration, *initiation* involves recognizing opportunities to replicate and deciding to act on them, while *initial implementation* is a process of “learning before doing.” As forms of exploitations, *ramp*
up to satisfactory performance is a process of learning by resolving unexpected outcomes, while integration involves maintaining and improving performance after satisfactory results are initially obtained.

25. Zollo and Winter (2002) are clear that this is an analytical representation and that, in practice, the processes of knowledge evolution described here are likely concurrent and confounded.

26. In this sense, the evolutionary logic can be understood as a broader learning strategy, of which such improvement-focused evaluation strategies as design-based implementation research are a core component.

27. Our framework can be understood as a theory-based approach to developmental evaluation (Patton, 2011; C. H. Weiss, 1997), though etic (i.e., grounded in the evolutionary logic) rather than emic (i.e., grounded in the network’s own theory of action). Moreover, it is a reciprocal approach. Specifically, it is an external perspective on the internal structure and operation of the network, with two interdependent purposes: using the evolutionary logic (and associated evaluation framework) to motivate and inform critical analysis of the network as a learning enterprise, and using that experience to reflect critically on the logic (and the associated framework).

28. While derived from the evolutionary logic, our view is that these questions have face validity independent of the evolutionary logic, and potentially useful for analyzing enterprises smaller in scope than schoolwide improvement enterprises. We also recognize that 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 for networkwide improvement.


30. A shell enterprise can operate as part of a good faith effort to improve schools: for example, efforts in which a hub establishes a small set of similarly structured templates to support initial, exploratory, cross-template learning. A shell enterprise can also operate absent good faith efforts to improve schools: for example, when the hub seeks to capitalize on fees to schools and, thus, does not engage in costly efforts to develop capabilities; and/or when a newly adopting school seeks to capitalize on the reputational assets of networks to establish identity and secure legitimacy, though absent a commitment to reform core capabilities. See the discussion of “faux replication” in Baden-Fuller and Winter (2012) and Winter and Szulanski (2001, 2002) for discussion of differing motivations for what we describe as a shell enterprise.

31. Evaluation use is a topic that has received formidable attention from researchers, policy makers, funders, and others for decades, a key theme being difficulty in
making effective use of evaluation processes and results. As argued by Preskill and Torres (2000), evaluation that seeks to support transformative learning in organizational contexts (such as developmental evaluation as proposed here) requires communal, collaborative, dialogical approaches that are carefully structured and expertly supported. Next steps in our research agenda include (a) reviewing the literature on evaluation use through the lens of developmental evaluation as proposed here and (b) experimenting with protocols and guidance to support the type of communal, collaborative, dialogical approaches described by Preskill and Torres.

32. For example, see Peurach (2011) and Peurach and Glazer (2012) on efforts by Success for All to field a schoolwide model focused on K-6 reading: on one half of one core content area that enjoys some consensus regarding its basic knowledge base (as evidenced by the efforts of the National Reading Panel) and for which the What Works Clearinghouse and Best Evidence Encyclopedia suggest a body of possible, vetted components. Yet, even in this comparatively established domain, Success for All evolved consistent with the evolutionary logic reported here.

33. Research does provide an account of a large-scale educational reform effort reorganizing in ways consistent with the evolutionary enterprise: the reorganization of National Alliance for Restructuring Education as America’s Choice (Cohen, Peurach, Glazer, Gates, & Goldin, 2014; Glazer, 2009a, 2009b). However, this reorganization was driven by issues other than analysis of weaknesses in existing strategies for managing intellectual capital.


References


Author Biographies

Donald J. Peurach is an Assistant Professor of educational leadership and policy in the School of Education at the University of Michigan. His research focuses on large scale, practice-focused improvement efforts in underperforming schools.

Joshua L. Glazer is a Visiting Associate Professor of educational administration at the Graduate School of Education and Human Development at George Washington University. He is currently directing a study of Tennessee’s Achievement School District, an ambitious initiative to improve the State’s most under-performing schools.

Sarah Winchell Lenhoff is the director of policy and research at Education Trust-Midwest. Her work focuses on the intersections of policy and practice in school and educator improvement efforts.