

Formative Evaluation and Complex Health Improvement Initiatives: A Learning System to Improve Theory, Implementation, Support, and Evaluation

American Journal of Evaluation
2020, Vol. 41(1) 89-106
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1098214019868022
journals.sagepub.com/home/aje



V. C. Scott¹ , K. Alia², J. Scaccia³, R. Ramaswamy⁴,
S. Saha⁵, L. Leviton⁶, and A. Wandersman⁷

Abstract

Sustainable community health improvement often requires the implementation of complex interventions in complex systems. Drawing from the *Four Keys to Success* frame (theory, implementation, support, and evaluation), this article describes how we used a formative evaluation approach to foster a learning system capable of monitoring and addressing emerging community needs within the Spreading Community Accelerators Through Learning and Evaluation (SCALE) initiative—a national capacity-building effort to support 24 community coalitions' progress toward a Culture of Health. The formative evaluation approach resulted in critical advancements to the theory, implementation, and nature of supports provided in SCALE. These improvements enabled the SCALE evaluation team to shift from the initial focus on program implementation issues to a greater emphasis on downstream factors (community-level outcomes). The ability of formative evaluation to grapple with the emerging challenges of implementing complex interventions in complex systems makes it particularly valuable for community health improvement initiatives.

Keywords

formative evaluation, community health, program development, systems evaluation

¹ Department of Psychological Science, The University of North Carolina at Charlotte, Charlotte, NC, USA

² Chief Executive Officer, Serve & Connect, Columbia, SC, USA

³ Chief Operating Officer, The Wandersman Center, Columbia, SC, USA

⁴ University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

⁵ Institute for Healthcare Improvement, Boston, MA, USA

⁶ Robert Wood Johnson Foundation, Princeton, NJ, USA

⁷ Barnwell College, University of South Carolina, Columbia, SC, USA

Corresponding Author:

V. C. Scott, Department of Psychological Science, The University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC 28223, USA.

Email: victoria.scott@uncc.edu

Introduction

Most of us have experienced the unease of implementing an initiative before testing was completed, in hopes of making changes in real time as new information and feedback come in. When the initiative is about community health improvement, and complex support and delivery are part of the model, the challenges are compounded. The evaluation design must parallel and capture the complexity but be manageable enough that stakeholders can understand and engage with it.

Tom Chapel, Chief Evaluation Officer, Centers for Disease Control and Prevention (November 29, 2017)

Many health improvement initiatives attempt to bring new programs, processes, policies, or practices into organizations or communities. Some are large-scale and ambitious, intervening at the population level and targeting macro-level transformation (e.g., Aligning Forces for Quality, Healthy People 2020, Culture of Health (Painter & Lavizzo-Mourey, 2008; Plough, 2015, U.S. Department of Health and Human Services, 2018). Others are locally focused, involving a single organization (e.g., integrating behavioral health in a primary care setting; Scott et al., 2017) or a multisector collaboration within a community (e.g., Siegel, Erickson, Milstein, & Pritchard, 2018). The potential for systems improvement or transformation can excite stakeholders. However, this excitement is often tempered by the challenges of implementing strategies. The reality is that implementation of system-level health improvement initiatives is difficult and commonly leads to disappointing results, even when collaborations involve the “best and brightest” agents of change. For example, a review of 48 community change initiatives found that while most communities experienced some improvements at a program, organizational, or neighborhood level, the majority of these initiatives did not result in the targeted degree of community health improvement (Kubisch, Auspos, Brown, Buck, & Dewar, 2011). In fact, few efforts resulted in policy or systems-level reform and even fewer demonstrated substantive changes in population health (Kubisch et al., 2011).

Why is it that so many health improvement initiatives fall short of achieving targeted goals? And, perhaps more importantly, what can we do about it? We, the authors, have devoted decades of research to piloting, evaluating, and improving bold community change efforts with audacious goals often requiring fundamental systemic shifts. These initiatives occur in complex settings, involve complex interventions, and are often linked to deep-rooted social determinants. As such, they require tremendous collective energy, resources, and persistence to achieve targeted outcomes. While we still have much to learn, our work has taught us that the probability of achieving and sustaining outcomes can be increased by utilizing a rapid learning system to systematically prepare, assess, and improve the delivery of complex interventions in complex settings. This rapid learning system can be created through a formative evaluation approach.

Formative evaluation is a rigorous assessment method used to identify potential and actual influences on the progress and effectiveness of implementation (Stetler et al., 2006). This type of evaluation is particularly valuable during the stages of developing and implementing new methods, practices, policies, or procedures and for rapid-cycle testing of innovative approaches. This article describes how we used a formative evaluation approach to develop a learning system for the Spreading Community Accelerators Through Learning and Evaluation (SCALE) initiative, a large community health capacity-building effort aimed at fostering a national Culture of Health (Plough, 2015). As a lens for our evaluation approach, we draw upon the *Four Keys to Success* (theory, implementation, support, and evaluation; Wandersman, 2009), a heuristic for examining the life span of social improvement programs. The four keys represent four critical areas of focus when conducting formative evaluations of an intervention program. Our aim is to illustrate specific ways that formative evaluation can help navigate the complexity of community health improvement interventions and settings. In this article, we introduce an array of heuristics and frameworks that

Table 1. Summary of Key Evaluation Heuristics, Frameworks, and Approaches Guiding Spreading Community Accelerators Through Learning and Evaluation and Value to Complex Interventions.

Heuristic/Framework/Approach	Description	Value to Complex Interventions
Formative evaluation	A rigorous assessment process aimed at improvement and designed to identify potential and actual influences on the progress and effectiveness of implementation.	A formative evaluation approach identifies emergent needs. It can contribute to the development of a learning system that is equipped to respond to the issues of complexity that characterize large-scale improvement efforts. For example, issues of interdependence, intervening across multiple system levels, and managing diverse stakeholder interests.
Four keys to success	A heuristic for examining the life span of social improvement programs. The four keys are theory, implementation, supports, and evaluation.	The four keys offer a frame for evaluating and reflecting on large-scale initiatives. They are useful for thinking about interdependencies and alignment across key elements of an initiative.
Evidence-based system for innovation support (EBSIS)	An implementation science framework capturing four support system components for capacity building (tools, training, technical assistance, and quality assurance/quality improvement.)	Complex interventions generally involve different types of supports (tools, training, technical assistance, quality assurance/quality improvement). Each support has its own empirical basis and theory of change. EBSIS provides a frame for systematically assessing common support system components.
Inquiry–Observation–Reflection (IOR) Framework	A structured, multiprong method of data collection involving inquiry, observation, and reflection.	The range of issues salient across community settings and capacity to address those issues can be quite diverse in large-scale initiatives. The IOR Framework provides a backbone structure for fitting evaluation methods to both the evaluation question and characteristics of the community. This can help evaluators think systematically about the specific evaluation method(s) to deploy across facets of a complex intervention.
Quadruple helix of accountability	References four key stakeholder groups that share accountability in community health improvement initiatives: community members, implementation team, funding staff, evaluators.	In complex interventions, close collaboration among key stakeholder groups improves coordination, facilitates alignment across interests, and promotes joint accountability for initiative outcomes.

have guided our work; a summary of these appear in Table 1. The formative evaluation approach described in this article was recognized as exemplary through the *American Evaluation Association's 2017 Outstanding Evaluation Award*.

Complexity in Improving Community Health

The health and well-being of individuals are shaped by the economic, structural, and institutional circumstances in which they live—environmental factors that interact in unique and context-specific

ways. Interventions to improve health outcomes must attend to present-day and historical contexts. This was underscored by former Secretary General Ban-Ki Moon in 2009:

Deep inequities in health outcomes—the unfair and avoidable differences in health status seen within and between countries—persist. . . . Although some of the inequities in health outcomes are due to differences in access to health services, the majority is attributable to the conditions in which people are born, grow, live, work, and age. In turn, poor and unequal living conditions are largely the result of poor social policies and programs, unfair economic arrangements, and politics driven by narrow interests. (Syme & Ritterman, 2009, p. 3)

Given the interplay of social, political, and environmental factors in our everyday life, it is improbable that a single intervention will result in significant change in the health of communities. In fact, interventions to improve community health are, by definition, *complex interventions*: They consist of multiple components that act both independently and interdependently (Campbell et al., 2000). Further, these interventions are implemented in settings that are *complex systems*, defined as enclosed contexts characterized by unpredictable and nonlinear properties (Begun, Zimmerman, & Dooley, 2003; Finegood, Karanfil, & Matteson, 2008; Snowden & Boone, 2007). Successfully implementing complex interventions is contingent on stakeholders' ability to sense and respond to issues that were not anticipated or understood before implementation or that emerged during implementation. The intervention must be flexible enough to evolve as the system (local context and needs) changes. This makes the evaluation of these interventions particularly challenging.

Evaluating Interventions in Complex Settings

In the 1990s, design research was developed in the field of education in response to the need to study phenomena in the real world rather than in controlled settings (Brown, 1992; Collins, 1992). Design research involves developing formative experiments to test and refine educational designs that are based on principles from prior research (Collins, Joseph, & Bielaczyc, 2004). It aims to identify all the variables of interest that affect dependent variables. Unlike hypothesis testing, which targets a limited number of variables, design research examines all aspects of an intervention in order to develop a profile of the intervention in practice. Ongoing refinements are made until implementation is successful, with evaluation involving both the researchers and implementation participants. Over time, this method leads to the discovery of more robust interventions and provides insight into how interventions evolve.

Formative evaluation embodies the essence and objectives of design research. In formative evaluation, data are collected during implementation and used in a timely fashion to identify and address emerging challenges. The central aim is continuous, midcourse improvement (Rossi, Lipsey, & Henry, 2018). Modifications to the theory of change or to the implementation plan are welcomed in the interest of achieving and sustaining targeted health outcomes (Øvretveit, Leviton, & Parry, 2011).

Four Keys to Achieving Successful Outcomes

Wandersman (2009) examined a series of well-recognized social improvement efforts and extracted four keys to successful system change: *theory*, *implementation*, *supports*, and *evaluation*. These four keys reflect four critical components of an intervention program and provide a useful framework for evaluating community health improvement initiatives. We summarize the four keys here.

Theory. A theory provides a plausible model of causal mechanisms. A theory of change clarifies assumptions about the process through which change is expected to occur, and it specifies how both early and intermediate outcomes will pertain to achieving desired long-term change. Contextual

elements related to the setting and population can be incorporated into the theory of change. For complex systems, the theory of change can serve as the initial groundwork for the intervention plan. The theory may evolve as implementation proceeds.

Implementation. Implementation is “a specified set of activities designed to put into practice an activity or program of known dimensions” (Fixen, Naoom, Blasé, & Friedman, 2005, p. 5). Successful implementation requires attending to a variety of aspects including fidelity, dosage/intensity, quality, participant responsiveness, program differentiation, program reach, and adaptation (Dane & Schneider, 1998; Meyers, Durlak, & Wandersman, 2012). Implementation quality is particularly critical to achieving outcomes (Durlak & DuPre, 2008). In complex interventions, implementation components often occur at multiple ecological levels (e.g., individual, microsystem, organizational, community).

System supports. System supports refer to tangible and intangible resources that help build delivery system capacities for implementation and service provision. The *Evidence-based System for Innovation Support (EBSIS)* details four types of supports: tools, training, technical assistance/coaching, and quality assurance/quality improvement (Wandersman, Chien, & Katz, 2012). Supports in complex systems need to be flexible to effectively respond to changes in theory and implementation strategy.

Evaluation. Generally, evaluation has been used to assess fidelity to the theory of change (*Is the program being implemented as planned?*) and effectiveness of the program (*Did the program achieve results?*). For complex systems, evaluation needs to determine: *What worked for whom and when? And, how did the system change as a result?* In these circumstances, a formative evaluation approach that provides rapid-cycle information is needed to facilitate continuous learning and to inform midcourse adaptations. The Cynefin framework (Snowden & Boone, 2007; Van Beurden, Kia, Zask, Dietrich, & Rose, 2011) recommends a “probe-sense-respond” decision-making strategy for complex systems whereby some initial experimentation provides enough knowledge about the system to formulate an adequate response (Snowden & Boone, 2007). A formative evaluation approach supports this strategy.

In sum, the four keys to success require (1) a sound theory, (2) quality implementation, (3) sufficient support system elements, and (4) a practical evaluation. The Four Keys to Success provide a framework for understanding and responding to the continuous exchange between complex interventions and the settings in which interventions occur. While presented individually above, the four keys are interdependent. For instance, a health improvement initiative may begin with a sound theory of change, but issues can emerge during implementation that require a redefinition of the theory. In turn, a redesign of the support system and development of alternative implementation strategies become necessary. As changes occur, formative evaluation can provide rapid-cycle data to facilitate learning and continuous quality improvement. In addition, adaptations to the evaluation strategy might be required as new issues emerge. In our work with SCALE, we found that structuring our evaluation across these four keys, concurrently and iteratively, helped us gain insights that facilitated adaptive evolution of the initiative.

The SCALE Initiative

100 Million Healthier Lives’ SCALE is a national capacity-building initiative designed to capitalize on progress that communities have already made and to further their journey toward a Culture of Health—A vision of the Robert Wood Johnson Foundation (RWJF) for all individuals to have the opportunity to live a healthier life (see <https://www.rwjf.org/en/cultureofhealth/about.html>, for

additional information about the Culture of Health). The initiative is funded by the RWJF and led by the Institute for Healthcare Improvement, along with key community partners, including Communities Joined in Action, Community Solutions, and the Network for Regional Healthcare Improvement. SCALE seeks to build the capacities of community change-agents to create sustainable change within their communities. It also aims to build a network of support among these communities. Participating community change-agents live in the target communities and have extensive and diverse knowledge about their communities.

SCALE began with three broad goals:

- (1) Develop leadership capabilities within communities to improve health, well-being, and equity (on the journey to a Culture of Health).
- (2) Create vibrant relationships, functional networks, and teamwork between communities.
- (3) Create and improve an intercommunity system capable of sharing and spreading best practices among communities.

Guiding SCALE's theory of change (see Figure 1), these three goals were established based on insights from six initiatives that have achieved sustainable improvements in complex settings (Stout, Howard, Lewis, McPherson, & Schall, 2017). The Institute for Healthcare Improvement convened organizational leaders from several of these key initiatives to design an initiative that could support continuous improvements in population health, well-being, and equity across diverse communities (Stout et al., 2017).

The first phase of SCALE (SCALE 1.0) was implemented between 2015 and 2017. It included 24 community coalitions with diverse needs, resources, capacities, and health improvement goals. Each community coalition had at least three organizations from the community, resulting in a nested model of systems (multiple organizations nested in each coalition). Launched in 2017, SCALE 2.0 aims to further the progress of SCALE 1.0 communities and to spread insights about community transformation with new communities. This article focuses primarily on insights from SCALE 1.0. The full SCALE 1.0 Evaluation Report by Wandersman and colleagues (2017) is available online: <https://www.100mlives.org/wp-content/uploads/2018/01/74284GPFNR.pdf>

Elements of the SCALE intervention. A multicomponent support system that aligned with SCALE's theory of change was provided to the SCALE communities. These components are summarized below:

- (1) *Community Health Improvement Leadership Academy (CHILA) trainings* were multiday, in-person events designed to foster relationships between communities, build skills in community health improvement (e.g., leadership, program implementation and evaluation, quality improvement methods), and facilitate learning across communities. Training goals for each CHILA were informed by community participants via need assessments and training evaluation feedback. Community members were also directly involved in the design of CHILA trainings, providing input, and leading select training sessions. Four CHILAs were held during SCALE 1.0. Additional information about the evaluation of CHILAs is available in Hayes et al. (2016).
- (2) *Monthly coaching* was tailored to each community and offered between CHILAs. Coaches were teams of technical assistance providers affiliated with the four organizations leading SCALE (Institute for Healthcare Improvement, Communities Joined in Action, Community Solutions, Network for Regional Healthcare Improvement). Coaching sessions were pre-scheduled and available per request in a 1:1 or small group format. Coaching sessions both reinforced content taught during CHILAs and addressed emerging community issues.

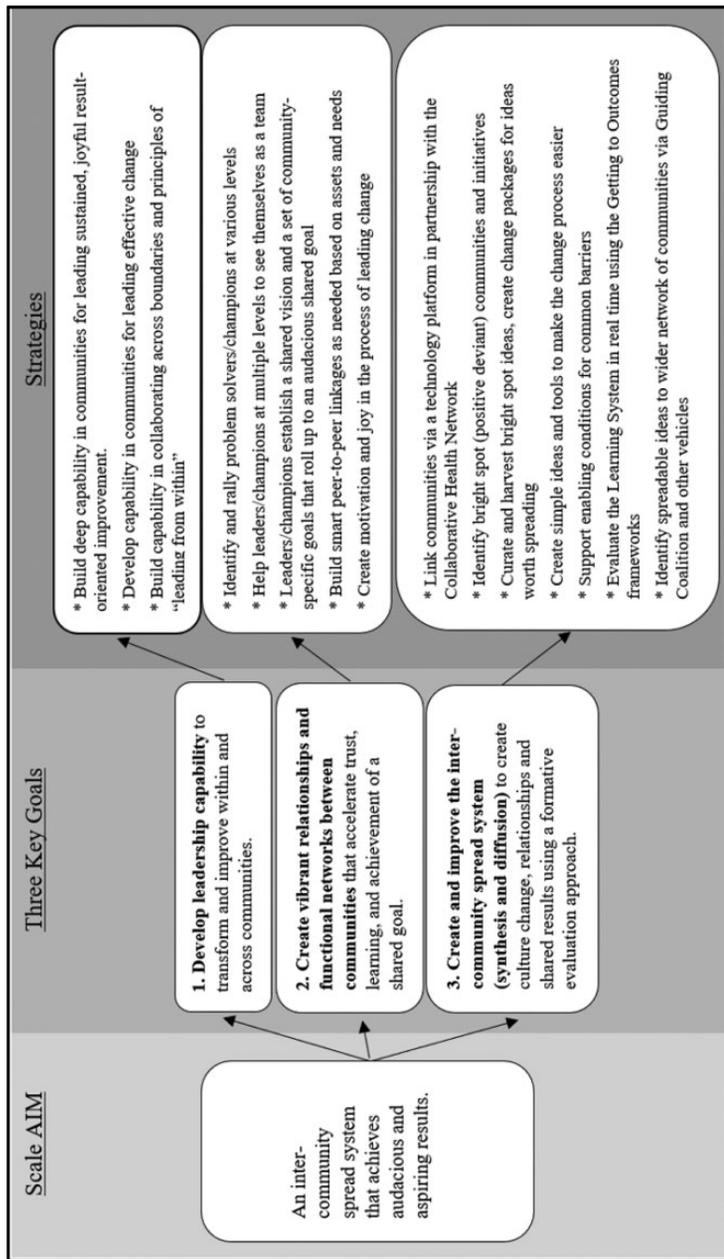


Figure 1. Spreading Community Accelerators Through Learning and Evaluation theory of change.

- (3) At the first CHILA, each community was selected into a *Peer Community Team (PCT)* based on community readiness for SCALE. Each PCT consisted of six communities and up to two coaches. Akin to a community of practice, PCTs facilitated peer-driven learning. Community members shared how they were applying SCALE methods to their own local community health projects; exchanged stories of their journeys toward community health improvement; provided encouragement to one another; and shared resources, practical tips, and lessons learned. This interteam opportunity was reported as one of the most valuable aspects of SCALE.
- (4) *Monthly webinars* were hosted by the SCALE implementation team to provide additional information and resources to community participants. Webinar topics were informed by community member needs and interests. As SCALE matured and the gifts of community members surfaced, community members began to serve as presentation leads on the webinars.
- (5) Community participants were connected to an *online social media platform* with a message board interface developed for members of the 100 Million Healthier Lives initiative. Monitored by SCALE implementation staff, the platform provided an asynchronous method for SCALE participants to access and offer both information and peer support. Data on participant use of the platform (e.g., use frequency and purpose) were incorporated in the SCALE evaluation.

Conceptualized as the critical facilitator of community improvement, the CHILAs were the key SCALE capacity-building events. *Action Periods* were designated after each CHILA. During the Action Periods, communities participated in the other SCALE support system components (coaching, PCT, webinars, social media platform) and worked with their coalition to apply the knowledge, skills, and resources acquired from the CHILA.

The Complexity of SCALE

The complexity of the SCALE initiative was defined by five attributes:

- (1) *A nested model of system complexity.* SCALE 1.0 involved 24 community coalitions with a minimum of three organizations per community. Each community-based organization operated independently as a complex system. With one designated convening organization serving as fiscal agent, organizations partnered on local community goals to form a coalition. Each community coalition became a complex system of its own. The 24 community coalitions were part of a larger complex system that also included SCALE's implementation team, evaluation team, and funders.
- (2) *Interdependence.* SCALE stakeholders worked closely together, sharing resources, ideas, and insights. The success of the initiative required stakeholders to consider existing interdependencies across multiple levels and stakeholder groups, including relationships among community-based organizations, across coalitions, and among community members, the implementation team, evaluation team, and the funders. The initiative also involved interdependence of key initiative change components (theory, implementation strategy, support system elements, and evaluation methods). Changes in any key components had implications for other components (e.g., change to theory required changes to the implementation and evaluation plans), as illustrated in the Results section.
- (3) *Theory of change with a multisupport system.* The SCALE intervention leveraged the EBSIS framework (Wandersman et al., 2012; see Table 1) and used multiple methods (e.g., tools, training, technical assistance, communities of practice, quality assurance/quality monitoring) for developing the capacity of its community change-agents. Each support

Table 2. Key Spreading Community Accelerators Through Learning and Evaluation Formative Evaluation Principles.

-
- Participatory approach with joint development of evaluation questions by key stakeholders (community members, implementation team).
 - Adaptation of the evaluation questions and evaluation methods as the theory of change evolves.
 - Real-time feedback to community members and implementation team members.
 - Multiple methods for assessing multiple stakeholder needs (e.g., inquiry, observation, reflection).
 - Multifaceted instead of dichotomous outcomes (e.g., *What worked well? What did not work well? Why did things work or not work well? What actions were taken when things didn't work well?* rather than *Did it work or not work?*).
-

method had its own theory of change and had influence on the overall SCALE theory of change. For example, trainings were designed to introduce and reinforce quality adoption of SCALE tools. Additionally, subsequent to each CHILA training, technical assistance/coaching was paired with ongoing evaluation to facilitate rapid-cycle learning and improvement.

- (4) *Multiple actors with diverse and changing community capabilities, needs, and goals.* Each individual participant of SCALE had his/her/their own set of competencies, needs, and interests. Effective collaborative decision-making required participants to recognize and negotiate variations and similarities in participant characteristics. Additionally, the SCALE team had to be responsive to the dynamic nature of partnering and adjust to the natural turnover in stakeholder participation over time. For example, between CHILA 1 and CHILA 4, turnover in participation per organization ranged from 0% to 100%, with an average of 59%. The use of a multisystem of supports helped the SCALE team respond efficiently and appropriately to changes in participating community members by tailoring conversations, resources, and the pace of information/resource sharing to changes in community needs.
- (5) *Dynamic environment.* Each community setting varied in levels of volatility, ambiguity, and uncertainty. Community challenges and solutions were not always apparent. Unclear causal linkages and confounding issues also existed within each setting as communities were simultaneously engaged in other community health improvement efforts. These setting characteristics brought additional challenges to the objective of measuring and monitoring the effectiveness of SCALE-specific interventions.

Method: Using Formative Evaluation in SCALE

With the Institute for Healthcare Improvement's specialization in rapid-cycle improvement in complex settings and the evaluation team's specialization in formative evaluation, SCALE provided an opportunity to test the utility of nimble evaluation methods for improving complex interventions occurring in complex settings. A mixed-method, multilevel formative evaluation strategy was uniquely developed for SCALE 1.0. The formative evaluation approach was informed by three evaluation approaches designed to provide iterative feedback to improve program implementation: Developmental Evaluation (Patton, 2011), Learning Evaluation (Balasubramanian et al., 2015), and FORECAST (FORMative Evaluation Consultation and Systems Technique; Katz et al., 2013). See Table 2 for principles that guided our evaluation.

The evaluation team involved nine core members who worked closely with the funder (RWJF), the implementation team, and community coalitions (program recipients) to design, implement, and

improve evaluation activities over the course of SCALE. The evaluation started with the following broad evaluation questions jointly developed by the evaluation team, the implementation team, and the funder:

- What accelerates or slows progress in SCALE?
- How do we use indicators of community readiness to monitor and predict progress?
- What helps spread good ideas between communities?
- How is spread influenced by other community initiatives?
- How do we make spread better and more efficient?

These broad questions guided the evaluation process, which involved the following iterative steps:

- (1) Periodic discussion of evaluation priorities with the funder (monthly) and the implementation team (at least weekly);
- (2) Development of evaluation methods to address the priorities (instruments, interview guides, participatory dialog, etc.);
- (3) Data collection, analysis, and synthesis;
- (4) Presentation and multistakeholder discussion of the results;
- (5) Recommendation of changes to one or more of the four keys (theory, implementation, supports, evaluation).

While core evaluation questions were jointly developed, there was not a rigid adherence to a set of evaluation questions or methods. Rather, the questions, methods, results, and actions evolved as the SCALE initiative progressed in order to remain sensitive to emerging issues. However, we emphasize that remaining open to emergence does not imply a lack of a systematic approach. Overall, the evaluation followed the SCALE theory of change and used a team-generated, multi-pronged method of inquiry, observation, and reflection referred to as the Inquiry–Observation–Reflection (IOR) Framework (see, e.g., Hayes et al., 2016). Both structure and flexibility were essential in the SCALE formative evaluation. The structure supported systematic and consistent data collection; the flexibility of our evaluation approach enabled adaptations to the program theory and the interventions. A logic model is presented in Figure 2 to illustrate the link of formative evaluation to SCALE’s complexity and the potential value of rapid-cycle evaluation use.

Results: Improvements Across the Four Keys Through Formative Evaluation

In this section, we provide examples of how formative evaluation facilitated SCALE program improvements across the four keys: theory, implementation, supports, and evaluation.

Evolution of Theory

During the second CHILA (multiday, in-person training), participant evaluation data stimulated SCALE stakeholder conversations about health equity in relation to SCALE’s theory of change. The SCALE team provided specific, health equity–focused modules during CHILA 2, as the topic of health equity was recognized as critical to community health improvement. However, SCALE community participant feedback indicated that the issue of health equity necessitated even more attention and resources. Specifically, they indicated that the topic should be interwoven throughout SCALE rather than addressed in a modular format. Open, collaborative discussions were had involving the implementation team and community members to give voice to the issue and to generate ideas for how best to enhance equity as a focus of SCALE. These conversations resulted

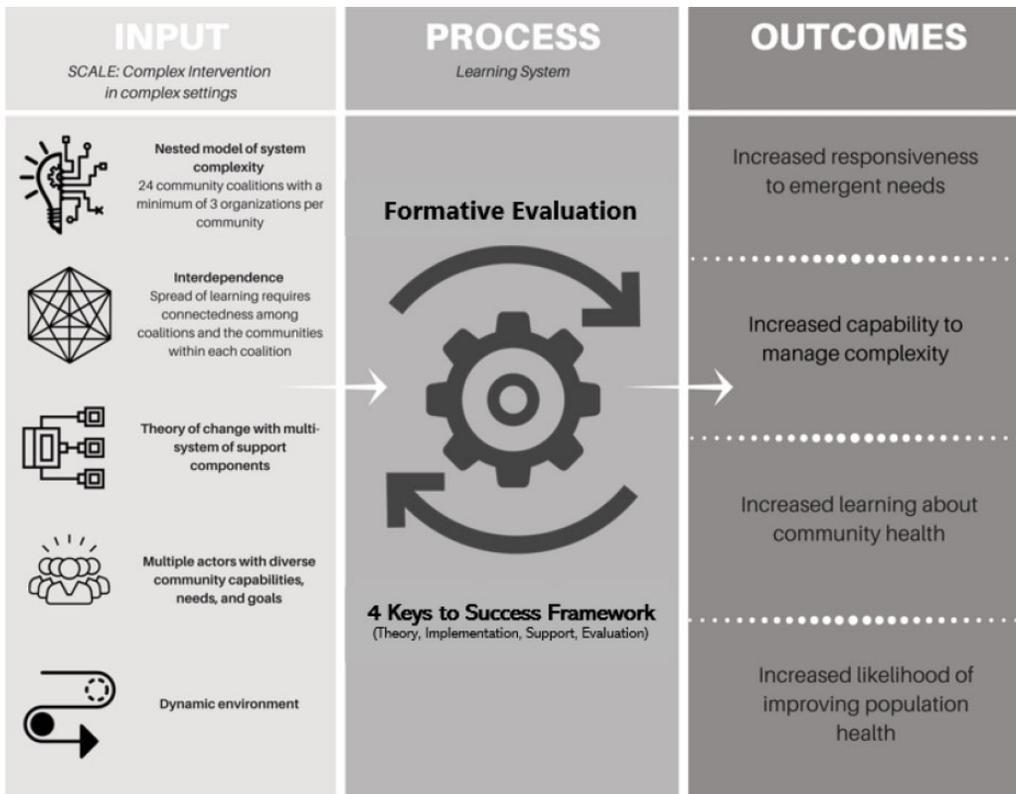


Figure 2. Representation of the Spreading Community Accelerators Through Learning and Evaluation formative evaluation approach.

in the decision to add health equity as a key driver in SCALE’s theory of change. This revision to the theory of change required subsequent changes to SCALE’s implementation plan, the content and design of support system components, and evaluation questions. For example, the content, keynote speakers, and tools presented at CHILA 3 all centered on how equity influences community health. The value of this shift for communities was shared by a community member with the evaluation team at the conclusion of a CHILA training:

I have taken a much greater interest in our capacity as a coalition and community to have courageous dialogue across barriers of cultural, racial, gender, socioeconomic, and ideological prejudice. This work is so vital to our community at this time . . . energizing other coalition members to see the connection between equity and public health would have been a tremendously difficult project without SCALE. I am also inspired by how well our community coalition has responded to the invitation to come to the next CHILA about equity to see how that might serve us in our work.

Evolution of Implementation

Process data collected on implementation quality revealed that particular core quality improvement activities (such as the use of Plan-Do-Study-Act [PDSA] cycles to test improvement ideas) were being implemented with inconsistent quality by community members. This is a common implementation issue, even in more controlled health-care settings (Taylor et al., 2013). Process

evaluation data were discussed with the implementation team. In response, a new capacity-building intervention (i.e., *Action Lab*) was introduced to improve implementation quality. The Action Lab added structure to the use of PDSA cycles by convening a group with the explicit power to enact and monitor desired changes. The process of continuous data collection and rapid feedback made it possible for the implementation team and community stakeholders to know when additional support system elements were necessary and the type of additional support to incorporate. The Action Lab model and more rigorous training on improvement methods were incorporated at the beginning of SCALE 2.0 to provide a robust structure for supporting improvement efforts.

Evolution of Support

SCALE webinars were initially designed as one-size-fits-all, with all participants receiving the same educational content. Administered in a web-based polling format, evaluations assessed the practical value of the session (e.g., *Please rate the degree to which the content will help you advance your work.*) and implementation factors (e.g., reach, participant engagement, quality). It also included open-ended questions to inform the design of subsequent webinars (e.g., *What suggestions do you have for future webinars?*). Evaluations revealed that webinars overemphasized theory and were too didactic. Community stakeholders desired a more interactive format. Moreover, the SCALE implementation team learned that the content of the webinars was not equally useful to community participants; participants varied widely in community needs and existing capacities. They expressed a desire for tailored webinars that would include time for peer-to-peer discussions about implementation issues. Per participant feedback, SCALE webinars were modified to include breakout groups, providing opportunities for community members to connect with peers as they reflected on their community experiences and the webinar curriculum. Subsequent evaluations indicated that this webinar change increased the perceived value of this support system component. Breakout groups were built into the design of SCALE 2.0 webinars.

Formative evaluations of the CHILAs informed not only advances in SCALE's theory of change (discussed in the *Evolution of Theory* section) but were also pivotal to improvements to the trainings. Drawing on the three-pronged IOR Framework, we used a combination of the evaluation methods (e.g., rapid feedback form, learning wall, questionnaire, observational form, critical moments reflection, and post-CHILA site visits; see Hayes et al., 2016, for detailed description of these training evaluation methods). For example, a 20' × 7' learning wall was placed in the training room to provide an open-ended opportunity for participants to post comments, questions, and suggestions and to read other participant messages throughout each training day. At the end of each training day, the SCALE implementation and evaluation teams reviewed posted comments. We reflected on suggestions for improvement, identified suggestions that could be immediately addressed, and discussed how improvements would be made. On the morning of the subsequent CHILA training day, participant suggestions and the SCALE implementation team responses were shared openly and discussed with training participants. We referred to this process of training improvement as "You Said.../We Did..." A few instances from CHILA 1 are included below:

You (community participants) said: "A lot of information was provided in a short period of time. It was difficult to navigate the tasks, the information, and requested activity timelines."

In response, we (SCALE implementation team)

- added orientation times to the beginning of each day and
- sequenced messages and training materials based on when the information was needed (e.g., present logistics information and agenda at the beginning of the day; circulate select handouts at the time of training modules rather than all at once in the mornings).

You (community participants) said: “Long days; need for time to absorb information.”

In response, we (SCALE implementation team)

- created an agenda with shorter days, more breakout sessions, more time for reflection, and optional activities and down time.

You (community participants) said: “Need more time to establish and build on relationships within and between communities.”

In response, we (SCALE implementation team)

- modified the design of training activities to facilitate interaction with existing and new community members.

These are just a few of many instances in which the use of rapid-cycle evaluation methods enabled the SCALE team to make real-time and long-term improvements to the support system.

Evolution of Evaluation

During SCALE 1.0, the evaluation team conducted routine meta-evaluations (evaluation of the evaluation) and made ongoing adjustments to the evaluation plan to optimize its relevance and value for SCALE stakeholders. Meta-evaluations used a combination of surveys, interviews, and technology-based analytics. Surveys solicited attitudinal and experiential feedback from implementation team and community members about the format and frequency of evaluations as well as the quality of engagement with the SCALE evaluation team. Sample Likert-based meta-evaluation items for community participants included: *Evaluation requirements are burdensome*, *The Evaluation Team members are responsive to our questions*, and *I feel valued and respected by the Evaluation Team*. Sample qualitative included: *What can the Evaluation Team do to further improve its relationship with SCALE communities?* and *What readiness survey items were less clear for you?* The evaluation team also examined the extent to which evaluation reports were viewed by community participants and actively used to drive decisions. Sample evaluation items administered to SCALE implementation team staff included: *Evaluation results are reported in a timely fashion*, *The evaluation approach is consistent with the aims of SCALE*, and *Evaluation reports are easy to read*.

Feedback indicated that the formative evaluations were perceived as relevant and useful. This critically allayed concerns that the SCALE evaluation team had about participant burnout associated with the frequency of evaluations.

The evaluation team also reflected regularly on the use of evaluation resources and which areas of SCALE to prioritize as the initiative evolved. Where appropriate, modifications were made to the evaluation plan. For instance, in the early stage of SCALE, the evaluation focused on the effectiveness of the support system components (CHILA, coaching, PCTs, webinar, social media platform). As implementation barriers associated with support systems components were resolved, attention was shifted to understanding the use and impact of SCALE resources within communities. Evaluation resources were shifted from monitoring the SCALE support system to evaluation of community-level implementation. This was a valuable, but unanticipated, shift. Using a rapid-cycle feedback approach helped to accelerate SCALE team learning and improvement about support system features. The early learning and adjustments enabled the SCALE team to turn their attention to downstream impact (community, rather than program-level, indicators), the ultimate interest of SCALE. The emphasis on community processes and outcomes was built into the initial evaluation plan for SCALE 2.0. It includes a focus on communities' implementation of improvement activities, community readiness for

engaging in health improvement efforts, and the process by which SCALE 2.0 communities spread the SCALE model to new communities.

A sample of common challenges to implementing complex interventions, and the utility of formative evaluation across the four keys, is available in Table 3.

Discussion: Reflections and Practical Insights for Using a Formative Evaluation Approach

The notion of complexity (complex challenges, complex interventions, complex systems) is a common concept in large health improvement initiatives. However, the reoccurring finding that community-based health improvement initiatives fall far short of targeted and transformative outcomes indicates that there is much to be learned in measuring and navigating complexity. Our SCALE experience demonstrated that a formative evaluation approach framed through the lens of the Four Keys to Success can effectively increase the probability of success for a new, bold initiative. It creates opportunities for experimentation, discovery, and learning while ensuring accountability in complex health improvement initiatives.

We share several practical insights from our experience that we believe can help increase the value of formative evaluations in other initiatives. First is the importance of joint accountability, in which all key stakeholders have a shared responsibility for the initiative's outcomes. In SCALE 1.0, we referred to this as the *quadruple helix of accountability*—comprised of the implementation staff, funding agency, community members, and evaluation team (see Figure 3). All aspects of the evaluation (from design to data collection and interpretation) were conducted in close collaboration with SCALE stakeholders. The following are a few examples of how SCALE team members collaborated:

- The SCALE evaluation team interacted with community participants from the outset and on a continuous basis. Evaluation findings were shared frequently with community members using multiple modalities (e.g., SCALE social media platform, in real time during CHILA trainings, via individualized community reports).
- Several evaluation team members participated on weekly implementation team calls, and implementation team members reciprocally joined evaluation team meetings.
- Evaluation and implementation team leadership met with staff from the funding agency on a monthly basis.
- The funding agency was accountable for linking the implementation and evaluation team to resources, building lessons learned from SCALE 1.0 into the development of SCALE 2.0, and good stewardship of philanthropy funds.

The SCALE implementation team viewed evaluation as central to the initiative. The collaborative approach to SCALE's evaluation enabled the evaluation to be continuously relevant and valuable.

Second is the importance of fostering a learning culture in which evaluation is perceived as an essential vehicle to continuous growth and improvement rather than a judgment of performance. During the onboarding stage of SCALE 1.0, community participants were oriented to SCALE requirements, which included conversations about the formative evaluation approach. The notion of "failing forward" was promptly introduced to encourage risk-taking and to emphasize the value of learning from testing and failure. This was a guiding principle for the implementation team and was shared with the communities as a key learning objective. The "improvement-oriented mind-set" depersonalized evaluation findings, reduced defensiveness, and increased SCALE stakeholder receptivity to participating in a robust evaluation effort.

Table 3. Examples of Practical Challenges, Common Pitfalls, and Application of Formative Evaluation Across the Four Keys to Success (Theory, Implementation, Supports, Evaluation).

Four Keys to Success	Practical Challenge	Common Pitfall	Formative Evaluation Applied
Theory	The theory of change for a bold initiative may not be fully conceptualized at project start resulting in ambiguous aspects. Upon establishing a theory of change, contextual demands may require the theory of change to be modified.	Being wedded to a theory of change and missing opportunities to realign efforts with contextual needs.	Routinely revisit theory of change with stakeholders (implementation team, funder, intervention recipients). Discuss implications of midcourse evaluation findings on the theory of change.
Implementation	Variable uptake of intervention components across communities.	Assuming that engagement and uptake of intervention components will be high, because community members are bright and highly motivated, and overlooking important contextual and motivational issues that impact capacity to engage in intervention components as a result.	Continuously monitor implementation attributes (e.g., dose, quality, participant responsiveness). Discuss emerging implementation issues with intervention stakeholders to facilitate improvement.
Support system	Delivering support system components (e.g., tools, training, technical assistance) that optimize cost-effectiveness and cost-efficiency.	Utilizing support system methods that are cost-efficient (economical), but not cost-effective (e.g., new knowledge, skills, and tools are difficult to put to use because they have been provided with inadequate consideration of the unique needs of stakeholders and circumstances of their organizations).	Conduct rapid-cycle, short evaluations of support system components (e.g., post-webinar polling; midcourse training evaluation). Midcourse evaluations increase opportunities to improve the practical value of supports. Using short (<5 min) assessments minimizes participant evaluation burden.
Evaluation	Stakeholders view evaluation as core (rather than peripheral) to an improvement initiative.	Designing an evaluation with minimal input from implementation team and community members.	Use a joint accountability approach to evaluation that involves working closely with implementation staff, community members, and funders. Also routinely conduct meta-evaluations (evaluation of the evaluation aim, questions, methods, capacities, etc).

Third, formative evaluations are time- and resource-intensive. This type of evaluation involves frequent points of data collection, rapid data analysis and turnaround, and continuous revisions to the evaluation plan. For large-scale initiatives, like SCALE, the intensity of effort and resources is



Figure 3. Quadruple helix of accountability.

magnified by the multitude of intervention components. Capturing, synthesizing, and providing feedback in rapid, short-cycles demands sustained vigilance to the intervention and stakeholders. Using a collaborative, joint accountability approach to evaluation extends the time required to conduct each evaluation stage. Recognizing, discussing, and preparing for these capacity requirements in the early stages of an initiative is recommended. We also found it helpful to simplify and automate data collection where possible (e.g., use short instruments and mobile data collection tools).

Conclusion

This article describes key elements and application of a learning system used to support a complex community health improvement initiative. The learning system leveraged a formative evaluation approach that (1) was guided by the Four Keys to Success framework, (2) enabled the SCALE team to continuously monitor diverse stakeholder needs, and (3) helped the implementation team to sense and timely respond to unanticipated, emergent issues and to engage in rapid-cycle testing of innovative solutions. A formative evaluation approach can be particularly valuable for initiatives where stakeholders are metaphorically “building the boat while it is sailing.” The ability to be responsive to evolving community needs elevated the progress of SCALE 1.0. Drawing upon our experience with SCALE, we view formative evaluations as a core component of complex improvement efforts that provide “win-win” benefits to communities, implementation teams, funders, and evaluators.

Author’s Note

A. Wandersman is now affiliated with Wandersman Center, Columbia, SC, USA.

Acknowledgments

The 100 Million Healthier Lives Spreading Community Accelerators Through Learning and Evaluation (SCALE) initiative was made possible with the support of the Robert Wood Johnson Foundation, and joint contributions of the SCALE Implementation and Evaluation teams, SCALE coaches, and SCALE community members. We are particularly grateful to the following colleagues who meaningfully contributed toward a

formative evaluation approach and generated so much joy among the SCALE teams: Laura Brennan, Hayley Browdy, Brittany Cook, Paul Howard, Niñon Lewis, Aleya Martin, Gareth Parry, Amy Reid, Marie Schall (SCALE implementation team project director), and Hanna Sherman.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by the Robert Wood Johnson Foundation.

ORCID iD

V. C. Scott  <https://orcid.org/0000-0002-8382-283X>

References

- Balasubramanian, B. A., Cohen, D. J., Davis, M. M., Gunn, R., Dickinson, L. M., Miller, W. L., . . . Strange, K. C. (2015). Learning evaluation: Blending quality improvement and implementation research methods to study healthcare innovations. *Implementation Science, 10*, 1–11.
- Begun, J. W., Zimmerman, B., & Dooley, K. J. (2003). Health care organizations as complex adaptive systems. In S. S. Mick & M. E. Wyttenback (Eds.), *Advances in health care organization theory* (pp. 253–288). San Francisco, CA: Jossey-Bass.
- Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions. *Journal of the Learning Sciences, 2*, 141–178.
- Campbell, M., Fitzpatrick, R., Haines, A., Kinmouth, A. L., Sandercock, P., Spiegelhalter, D., & Tyrer, P. (2000). A framework for design and evaluation of complex interventions to improve health. *British Medical Journal, 321*, 694–696.
- Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O’Shea (Eds.), *New directions in educational technology* (pp. 15–22). Berlin, Germany: Springer-Verlag.
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences, 13*, 15–42.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23–45.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350.
- Finegood, D. T., Karanfil, O., & Matteson, C. L. (2008). Getting from analysis to action: Framing obesity research, policy and practice with a solution-oriented complex systems lens. *Healthcare Papers, 9*, 36–41.
- Fixen, D. L., Naoom, S. F., Blasé, K. A., & Friedman, R. M. (2005). *Implementation research: A synthesis of the literature*. Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).
- Hayes, H., Scott, V., Abraczinskas, M., Scacciam, J., Stout, S., & Wandersman, A. (2016). A formative multi-method approach to evaluating training. *Evaluation and Program Planning, 58*, 199–207.
- Katz, J., Wandersman, A., Goodman, R. M., Griffin, S., Wilson, D. K., & Schillaci, M. (2013). Updating the FORECAST formative evaluation approach and some implications for ameliorating theory failure, implementation failure, and evaluation failure. *Evaluation and Program Planning, 39*, 42–50.
- Kubisch, A., Auspos, P., Brown, P., Buck, E., & Dewar, T. (2011). Voices from the field III: Lessons and challenges for foundations based on two decades of community-change efforts. *The Foundation Review, 3*, 138–149.

- Meyers, D. C., Durlak, J. A., & Wandersman, A. (2012). The quality implementation framework: A synthesis of critical steps in the implementation process. *American Journal of Community Psychology, 50*, 462–480.
- Øvretveit, J., Leviton, L., & Parry, G. (2011). Increasing the generalisability of improvement research with an improvement replication programme. *BMJ Quality and Safety, 20*, 87–91.
- Painter, M. W., & Lavizzo-Mourey, R. (2008). Aligning forces for quality: A program to improve health and health care in communities across the United States. *Health Affairs, 27*, 1461–1463.
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York, NY: Guilford Press.
- Plough, A. (2015). Building a culture of health: A critical role of public health systems and services research. *American Journal of Public Health, 105*, S150–S152.
- Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2018). *Evaluation: A systematic approach* (8th ed.). Thousand Oaks, CA: Sage.
- Scott, V. C., Kenworthy, T., Godly-Reynolds, E., Bastien, G., Scaccia, J., McMickens, C., . . . Wandersman, A. (2017). The Readiness for Integrated Care Questionnaire (RICQ): An instrument to assess readiness to integrate behavioral health and primary care. *American Journal of Orthopsychiatry, 87*, 520–530.
- Siegel, B., Erickson, J., Milstein, B., & Pritchard, K. E. (2018). Multisector partnerships need further development to fulfill aspirations for transforming regional health and well-being. *Health Affairs, 37*, 30–37.
- Snowden, D. J., & Boone, M. E. (2007). A leader's framework for decision making. *Harvard Business Review, 85*, 68–76.
- Stetler, C. B., Legro, M. W., Wallace, C. M., Bowman, C., Guihan, M., Hagedorn, H., . . . Smith, J. L. (2006). The role of formative evaluation in implementation research and the QUERI experience. *Journal of General Internal Medicine, 21*, S1–S8.
- Stout, S., Howard, P., Lewis, N., McPherson, M., & Schall, M. (2017). *Foundations of a community of solutions, SCALE 1.0 synthesis reports*. Cambridge, MA: Institute for Healthcare Improvement.
- Syme, S. L., & Ritterman, M. L. (2009). The importance of community development for health and well-being. *Community Development Investment Review, 3*, 1–13.
- Taylor, M. J., McNicholas, C., Nicolay, C., Darzi, A., Bell, D., & Reed, J. E. (2013). Systematic review of the application of the plan–do–study–act method to improve quality in healthcare. *BMJ Quality and Safety, 23*, 290–298.
- U.S. Department of Health and Human Services. (2018). *Healthy people 2020*. Retrieved from <https://www.healthypeople.gov/>
- Van Beurden, E. K., Kia, A. M., Zask, A., Dietrich, U., & Rose, L. (2011). Making sense in a complex landscape: How the Cynefin framework from complex adaptive systems theory can inform health promotion practice. *Health Promotion International, 28*, 73–83.
- Wandersman, A. (2009). Four keys to success (theory, implementation, evaluation, and resource/system support): High hopes and challenges in participation. *American Journal of Community Psychology, 43*, 3–21.
- Wandersman, A., Chien, V. H., & Katz, J. (2012). Toward an evidence-based system for innovation support for implementing innovations with quality: Tools, training, technical assistance, and quality assurance/quality improvement. *American Journal of Community Psychology, 50*, 445–459.
- Wandersman, A., Scaccia, J., Alia, K., Ramaswamy, R., Parry, G., Reid, A., . . . Cook, B. (2017). *Spreading community accelerators through learning and evaluation: Findings from the SCALE formative evaluation*. Retrieved from <https://www.100mlives.org/wp-content/uploads/2018/01/74284GPFNR.pdf>