SECTION 3, CHAPTER 10

HOW IMPLEMENTATION SCIENCE AND IMPROVEMENT SCIENCE CAN WORK TOGETHER TO IMPROVE EARLY CARE AND EDUCATION

Tamara G. Halle, Ph.D., Child Trends

Now is an exciting time in early childhood research as well as program and policy development. Researchers are using new and innovative methods to explore the effectiveness of early childhood programs and policies with different populations and in varying circumstances. Researchers and policymakers are greatly interested in determining what it takes to improve the quality of early care and education (ECE) and achieve the outcomes we want for young children, especially those from low-income backgrounds. Two new perspectives, implementation science and improvement science, are being brought to bear on these important questions. Implementation science is an interdisciplinary field, encompassing different scientific disciplines (e.g., behavioral psychology, behavioral economics, sociology), different occupations (e.g., administrators, frontline implementers, trainers, researchers), and different service sectors (e.g., education, health) (Øvretveit, n.d.). It aims to bridge the gap between evidence of effective interventions and what is done in practice. Implementation science research is relatively new and has mainly been carried out in the social service fields of health, mental health, child welfare, and education (Century & Cassata, 2016; Damschroder et al., 2009; Peters, Adam, Alonge, Agyepong, & Tran, 2013). Only recently has implementation science begun to be used in ECE (Halle, Metz, & Martinez-Beck, 2013), and this framework is still not widely understood among early childhood researchers or practitioners. However, because of its success in other sectors, interest is growing in incorporating an implementation science perspective into our investigations of what works in ECE, with the hope that such a perspective can help us uncover the distinct components of complex programs or systems that are associated with changes in outcomes (i.e., the "critical ingredients" of early childhood programs and systems), help practitioners achieve the goals of early childhood programs, and support taking effective ECE programs or systems to scale (Halle et al., 2013; Yoshikawa, Wuermli, Raikes, Kim, & Kabay, 2018).

At the same time, because of the strong focus on quality improvement (QI) in ECE programs and systems throughout the United States (Derrick-Mills, Sandstrom, Pettijohn, Fyffe, & Koulish, 2014; Schaack, Tarrant, Boller, & Tout, 2012; Tout, Epstein, Soli, & Lowe, 2015; Wesley & Buysse, 2010; Young, 2017), there is growing interest in the burgeoning field of improvement science and its promise to promote a culture of quality improvement in early childhood settings (Boller, Sciarrino, & Waller, 2018; Daily et al., 2018; Hetzner, Arbour, Douglass, Mackrain, & Agosti, 2018). Like implementation science, improvement science has been used extensively in health care (Grol, Baker, & Moss, 2002; Improvement Science Research Network, 2010; Institute for Healthcare Improvement [IHI], 2003). Improvement science uses foundational concepts developed in business and manufacturing (Deming, 1986) and also draws on implementation science, systems theory, behavioral science, and change management (Daily et al., 2018). It has expanded to disciplines including education, child trauma, and child welfare (Agosti, Conradi, Halladay Goldman, & Langan, 2013; Bryk, 2015; Ebert, Amaya-Jackson, Markiewicz, Kisiel, & Fairbank, 2012; Haine-Schlagel, Brookman-Frazee, Janis, & Gordon, 2013). Although QI initiatives in ECE are growing more common, how such initiatives are defined and implemented varies widely across ECE settings and systems (Daily et al., 2018; Derrick-Mills et al., 2014). Few early childhood researchers or ECE practitioners interested in quality improvement

are familiar with the systematic methods of improvement science. Furthermore, application of improvement science techniques in ECE and the study of this framework's effectiveness in ECE settings is just beginning (Arbour et al., 2016; Douglass, 2015).

Because implementation science and improvement science are new to the early childhood field, researchers may be confused about what taking an implementation science or improvement science perspective means when studying the effectiveness, adaptation, and/or scale-up of early childhood programs, policies, or practices. Furthermore, policymakers, practitioners, and researchers may struggle to understand how a study focused on implementation or quality improvement differs from what program evaluators have been doing for years when they study for whom and under what conditions early childhood programs and systems achieve their best results. In this chapter, I outline the commonalities and distinctions between implementation science and improvement science, and I demonstrate how they can enhance program development and program evaluation in early childhood settings. I contend that implementation science and improvement science, though distinct, share many common elements and are highly compatible. An understanding of what these different frameworks offer, in both their commonalities and their unique features, can support effectiveness and continuous improvement of programs, policies, and practices (hereafter referred to collectively as "interventions") in the early childhood field.

Implementation science and improvement science, though distinct, share many common elements and are highly compatible.

COMPARISON OF IMPLEMENTATION SCIENCE AND IMPROVEMENT SCIENCE

To compare implementation science and improvement science, it is best to consider what each framework claims as its core tenets and features.

Definitions and main aims

Implementation science is the systematic inquiry into the processes by which interventions are enacted in the real world. It examines not only the interventions themselves but also the contextual factors and organizational supports that are necessary to create a hospitable environment for enacted interventions to achieve their intended outcomes (Century & Cassata, 2016; Damschroder et al., 2009; Granger, Pokorny, & Taft, 2016; Martinez-Beck, 2013; Peters et al., 2013; Peters, Tran, & Adam, 2013). It typically focuses on the implementation of an evidence-based program or practice. Consequently, implementation science, like some program evaluations, is interested in *intervention fidelity*, that is, the extent to which the intervention was actually delivered "as designed" and intended (Hulleman, Rimm-Kaufman, & Abry, 2013). However, implementation science recognizes that evidence-based practices may need to be adapted to work in different contexts or for different individuals in new settings. Furthermore, implementation science can be used to explore innovations that have not yet been proven to be effective. Implementation science also focuses on *implementation fidelity*, that is, the extent to which the contextual, individual, and organizational supports for implementation of an evidence-based practice or an evidence-informed innovation are in place and functioning well (Hulleman et al., 2013). These core implementation supports include *implementation teams* (i.e., the individuals who are intentionally supporting implementation), the use of data and feedback loops in a recursive and iterative fashion to solve problems and improve practices, and *implementation infrastructure* (i.e., individual competencies, organizational processes and partnerships, and leadership) that support effective implementation (Fixsen, Blase, Duda, Naoom, & Wallace, 2009; Metz, Halle, Bartley, & Blasberg, 2013; Metz, Naoom, Halle, & Bartley, 2015).[1] Finally, implementation science emphasizes the need to address implementation supports throughout all *stages of implementation*, ranging from early exploration to full implementation and eventually sustainability (Aarons, Hurlburt, & Horowitz, 2011; Fixsen & Blase, 2008).

Improvement science involves a systematic examination of the methods and contextual factors that best facilitate quality improvement at the individual, program, and/or system level (Health Foundation, 2011; Langley et al., 2009; Shojania & Grimshaw, 2005). Improvement science draws heavily on process improvement models from business and manufacturing (Deming, 1986) and on organizational change management theory (Cameron & Green, 2009), as well as implementation science (Durlak & DuPre, 2008; Fixsen, Naoom, Blase, Friedman & Wallace, 2005; Meyers, Durlak, & Wandersman, 2012). Improvement science originated in manufacturing as the systematic study

[1] I cover these components of implementation infrastructure in more detail during the discussion of research questions later in this chapter.

of the series of steps and activities that make up a work process, with the aim of improving the quantity and/or quality of the work product and reducing costs. The inclusion of systems thinking and change management perspectives led to the study of how workers think together about improving their activities as a team. Improvement science strongly emphasizes the expertise of practitioners and their role as "active inquirers" who develop *practice-based evidence* (Bryk, 2015).

Two prominent methodologies that have come out of improvement science are the Breakthrough Series Collaborative (BSC; see IHI, 2003) and Collaborative Improvement and Innovation Networks (CoIINs; see Selk, Finnerty, Fitzgerald, Levesque, & Taylor, 2015).[2] Both of these methodologies share key features: they emphasize multidisciplinary, cross-role *collaborative teams*; they employ *expert faculty* or coaches who facilitate the collaborative teams within a *shared learning environment*; they explore evidence-based strategies to improve practices in a particular focal area; they make frequent and rapid *use of data* to test small changes, solve problems, and track progress using actionable metrics; and they promote changes in *organizational culture* as a way to keep the focus on learning and continuous quality improvement.[3] To instill a culture of learning and improvement, the emphasis tends to be on innovation and adaptation of practice to fit the current context rather than on fidelity to rigid standards of practice, which is often associated with a culture of compliance (Derrick-Mills et al., 2014).

Like implementation science, improvement science recognizes that evidence-based practices do not work the same way in all contexts or for all individuals. Professionals, therefore, need the freedom to make adaptations. But those adaptations must be systematically tested to ensure that they indeed improve outcomes (Taylor et al., 2014). A hallmark of improvement science is the use of Plan, Do, Study, Act cycles (PDSAs; see Deming, 1986) that let individuals determine, through the tracking of specific, actionable metrics, whether a small change in practice leads to improvements in outcomes. Improvement science also focuses on organizational capacity building through promotion of leadership at all levels of the organization (Conradi et al., 2011). Organizational capacity building is fostered by acknowledging the professionalism and expertise that all employees bring to the collaborative improvement process. The ability to build an organization's capacity and leadership for QI depends in large part

[2] Other improvement models, such as Lean, Six Sigma, Kaizan, Chronic Care Model, and Vermont Oxford Network have also been developed (Health Information Technology Research Center, 2013; Levinson & Rerick, 2002; Nadeem, Olin, Hill, Hoagwood, & Horwitz , 2013; Scoville & Little, 2014). BSC and CoIINs are the focus here because these two models have begun to be used in the early childhood field (Hetzner et al., 2018).

[3] In a CoIIN, the shared learning environment is sometimes virtual rather than face to face. This feature, and the duration of a CoIIN, are two of the few differences between the CoIIN and BSC models. In the BSC, the exploration of evidence-based strategies to improve practices in a particular focal area is referred to as the *change framework*. CoIINs have been applied to various focal areas; for example, they've been used to reduce infant mortality and to increase school readiness among children birth to age three. See https://www.nichq.org/impact/our-work/list for more. In the BSC, the frequent and rapid use of data to test is referred to as the *model for improvement*, which uses Deming's (1986) Plan, Do, Study, Act improvement process (Langley et al., 2009, p. 5; Scoville & Little, 2014, p. 6).

on the organization's culture. A culture that encourages risk-taking and a shared belief that making mistakes is part of the learning process provides a hospitable environment for growth and improvement. Improvement science claims that methodologies such as BSC or CoIIN help to accelerate learning, spread innovations, and improve both practice and outcomes faster than other methods such as one-on-one coaching (McPherson, Gloor, & Smith, 2015; Langley et al., 2009).

Looking across the definitions and aims of implementation science and improvement science, we see several commonalities. One is that they both highlight how the systematic study of practices can improve outcomes for individuals, programs, and/or systems as implemented in real-world conditions. A central aim of both implementation science and improvement science is bridging the gap between research and practice—that is, taking the evidence-based practices identified through rigorous program evaluation and studying how these practices are enacted in real-life settings (Ammerman, Putnam, Margolis & Van Ginkel, 2009; Tansella & Thornicroft, 2009; Wandersman et al., 2008). Both are also concerned with context and how that affects the success of an intervention, and both focus on identifying the mechanisms that support achieving improved outcomes.

A central aim of both implementation science and improvement science is bridging the gap between research and practice—that is, taking the evidence-based practices identified through rigorous program evaluation and studying how these practices are enacted in real-life settings.

What, then, distinguishes these frameworks? The distinctions are subtle. Implementation science tends to focus on the conditions that support *fidelity* to evidence-based or evidence-informed practices as a means to achieve the *intended outcomes* of an intervention, whereas improvement science does not (see Table 1). Rather, improvement science tends to focus on *innovation* and *adaptation* based on evidence-based practices as a means to achieve *improved outcomes*. However, implementation science also acknowledges and tests adaptations and is interested in improved outcomes, not just fidelity and intended outcomes (Century & Cassata, 2016). This may be why some researchers consider implementation research to be a type of improvement research (Olds et al., 2013).

Another difference is the time it may take to achieve outcomes. Implementation science posits that long-term outcomes may not be evident until full implementation of an evidence-based intervention has been achieved, which could take two to four years (Fixsen et al., 2005). In contrast, improvement science aims to make improvements in outcomes rapidly—for example, over the span of 12 to 18 months (McPherson et al., 2015). Evidence of sustainability of those improvements, however, is currently limited (Wells et al., 2017). A final distinction is that improvement science aims to develop practice-based evidence in addition to evidence-based practice (Bryk, 2015). In sum, in their main areas of focus, implementation science and improvement science appear to be more similar than different (see Table 1).

Table 1. Comparison of areas of focus and main aims for implementation science and improvement science

Areas of focus	Implementation Science	Improvement Science
Systematic study of practices to achieve improvements in outcomes	√	√
Local context	√	√
Real-world settings	√	√
Adaptation	√	√
Innovation	√	√
Intervention fidelity	√	
Implementation fidelity	√	√
Aims		
Bridging the gap between research (i.e., the evidence base) and practice	√	√
Developing the evidence base for evidence-based implementation practices	√	
Supporting and sustaining evidence-based practice outcomes	√	√
Building practice-based evidence		√
Achieving intended outcomes	√	
Achieving improved outcomes	√	√
Identifying mechanisms that support achieving improved outcomes	√	√
Identifying individuals for whom the intervention results in improved outcomes	√	√
Identifying the conditions under which improved outcomes are achieved	√	√
Achieving improvements in outcomes quickly		√

Key research questions

As with most evaluations and continuous improvement efforts, asking the right questions and getting them answered produces better outcomes.

Many of the research questions that traditional program evaluation examines are also of interest to implementation researchers. Specifically, implementation studies investigate the definition of what is being enacted in the real world (i.e., description of intervention components) and the description of processes by which an intervention is enacted and ask whether the intervention has been enacted as intended (i.e., intervention fidelity). Additionally, implementation research is interested in describing what adaptations, if any, were needed to ensure that the intervention's goals could be achieved in the current context.

Because implementation research is the study of how an intervention is enacted under real-world conditions, there is constant tension between measuring fidelity to a model and documenting adaptation or customization (Glasgow, 2009). Chambers, Glasgow, and Stange (2013) proposed an implementation model called the Dynamic Sustainability Framework to account for the changing contexts at both the level of the individual program and that of the broader ecological system within which interventions can be continuously refined and improved as they are sustained.

Since program evaluation and implementation research significantly overlap in what they typically address, implementation research is sometimes considered a type of program evaluation, one that focuses in particular on the processes of program implementation rather than participant outcomes.[4] However, implementation science also addresses questions that are not necessarily common in traditional program evaluation. For example, implementation science is more likely to emphasize documenting the role of implementation teams and the use of data and feedback loops (Metz et al., 2015). Like improvement science, implementation science emphasizes the importance of using data early and often (within iterative PDSA improvement cycles) to allow team members to adjust program components and/or implementation supports when initially developing an intervention, when implementing an evidence-based intervention in a new context, and/or when implementing at scale. Establishing data systems to continuously gather and use data is strongly encouraged as part of building the organizational infrastructure for effective implementation of an intervention. Researchers operating from an implementation science perspective will often ask the team members responsible for implementing the intervention what data they collect, how frequently they collect it, how they use the data they gather, and how the data are stored and analyzed.

[4] See, for example, the categories of program evaluation noted in the Fatherhood and Marriage Local Evaluation & Cross-site Project (http://www.famlecross-site.info/EvalDesign.html). I also discuss later in this chapter innovative evaluation designs, such as developmental evaluation, that embody implementation science principles.

Questions about data and feedback loops are related to another unique contribution of implementation science to program evaluation: the assessment of the existence, functioning, and quality of the implementation infrastructure to support an early childhood intervention model. Questions about implementation infrastructure focus on *staff competencies* (Do early childhood staff have sufficient knowledge of early childhood practices in general? What is the level of staff buy-in for this particular intervention model? Has staff been well trained in the intervention model?), *organizational processes* (What policies and practices are in place or are newly created that will support the intervention in this early childhood setting? What partnerships have been established or marshaled to support the intervention? How is information about the intervention's activities and outcomes collected, shared, and used by staff?), and *leadership* (Who is on the implementation team for this intervention in this early childhood setting? Is leadership represented at all levels of the organization and/or system? Are teachers and caregivers in early care and education settings viewed as leaders in implementing innovations? What do implementation team members do with the information about how the intervention is proceeding at this setting? How do leaders address the technical and adaptive challenges of implementation?). Specific implementation research questions also address the context in which implementation occurs as well as the individual, organizational, and systems capacity and readiness to take on new practices (Bumbarger, 2015; Peterson, 2013). In sum, implementation research questions often go one layer deeper than the general description of intervention processes and outcomes to identify the *who*, *what*, and *how* of successful implementation in real-world, practical contexts (Damschroder et al., 2009; Granger et al., 2016; Martinez-Beck, 2013; MEASURE Evaluation Working Group, 2012).

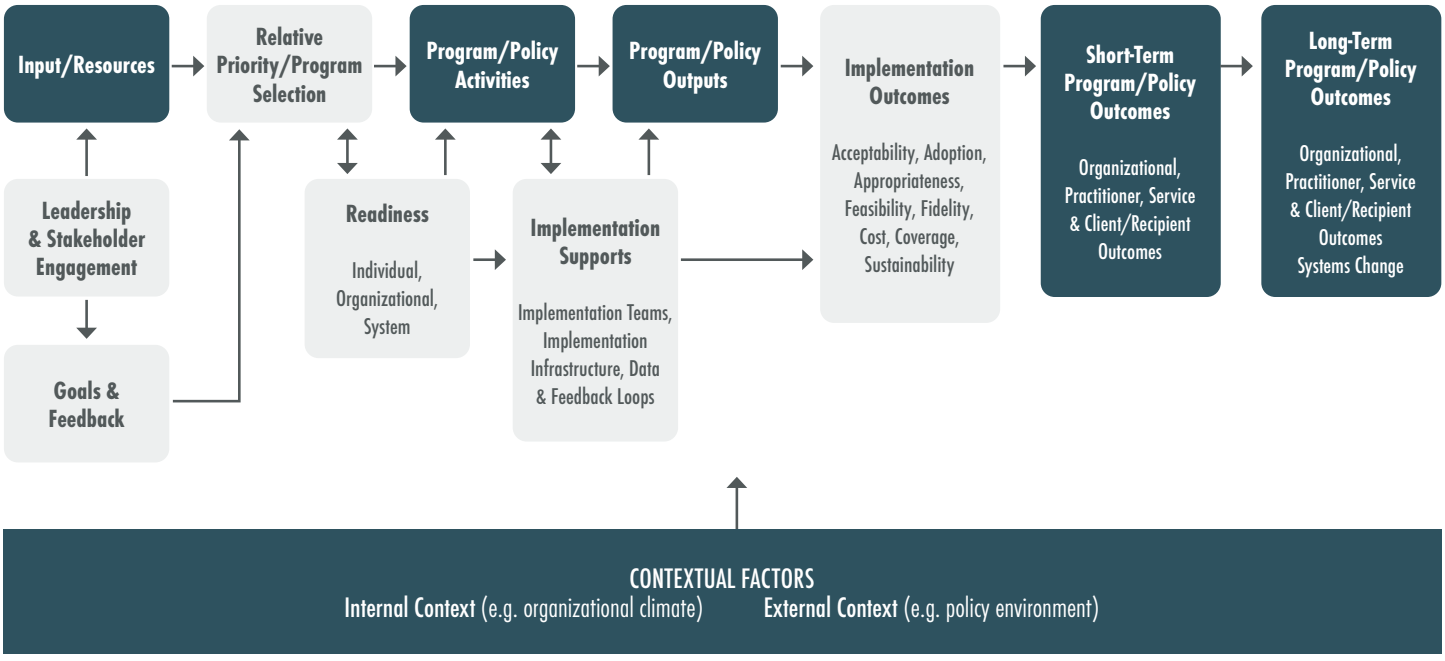
Another contribution that implementation science has made to traditional program evaluation is its treatment of implementation outcomes as distinct from intervention outcomes (Peters, Tran, & Adam, 2013; Proctor et al., 2011). Proctor and colleagues (2011) distinguished implementation outcomes from service outcomes, such as effectiveness and efficiency, and client outcomes, such as satisfaction. More recently, Peters and colleagues (Peters et al., 2013; Peters, Tran & Adam, 2013) adapted the implementation outcome variables proposed by Proctor and his collaborators so that they could be applied to both programs and policies. For example, specific implementation outcomes address questions about spread, scale-up, and sustainability (Century & Cassata, 2016).[5] Implementation science's unique contributions to program and policy evaluations are depicted in Figure 1, with implementation elements represented in gray and traditional program or policy evaluation components represented in blue.[6]

[5] Some researchers use the term *diffusion* to indicate what I am referring to as *spread* (Franks & Schroeder, 2013). Likewise, the terms *penetration* or *coverage* are sometimes used in lieu of *scale-up* (Peters et al., 2013; Proctor et al., 2011).

[6] Context is a central concern of implementation science, but it is also part of the logic model for most program evaluations. Therefore, I have depicted this element in blue.

Improvement science is particularly interested in empowerment and leadership at all levels of an organization as a means for instilling a culture of continuous improvement at individual, team, and organizational levels. Relatedly, improvement science documents the role of readiness in making changes at the individual, team, and organizational levels. Improvement science also asks questions about *organizational culture* and *climate* (Do the collective attitudes of those in this early childhood setting endorse a sense of psychological safety to make mistakes and try new things? Do these collective attitudes about the climate for supporting improvement change over time? What work processes and norms exist in this organization?) and the *spread* and *sustainability* of improvements (Are improvement activities, such as the use of data to test small changes in practice, being used by those outside of the initial group of individuals who had engaged in improvement activities? Are improvement practices being used in the early childhood setting to address improvement needs beyond the initial topic that was addressed by the improvement strategies?). Finally, improvement science is concerned with *explaining variability* in outcomes based on the interaction of organizational culture or norms and task requirements (Bryk, 2015). Although implementation science and improvement science overlap quite a bit in terms of research questions of interest (see Table 2), an emphasis on infusing a culture of inquiry and improvement in an organization and a deemphasis on fidelity to or compliance with particular practices are what most distinguish improvement science from implementation science (and also traditional program evaluation).

Figure 1. Conceptual model incorporating implementation elements into traditional program and policy evaluations.



Note: Incorporates concepts from Bauer et al. (2015), Brennan et al. (2013), Damschroder et al. (2009), Metz et al. (2015), and Proctor et al. (2011).

Table 2. Comparison of main research questions and outcomes of interest for implementation science and improvement science

Research Questions/ Outcomes of Interest	Implementation Science	Improvement Science
Acceptability	√	√
Adaptation	√	√
Adoption	√	√
Appropriateness/fit	√	√
Client outcomes	√	√
Cost	√	√
Dosage	√	√
Effectiveness	√	√
Equity	√	√
Feasibility	√	√
Feedback loops	√	√
Fidelity to intervention components	√	
Fidelity to implementation components	√	√
Implementation infrastructure	√	√
Implementation teams	√	√
Leadership	√	√
Needs	√	√
Organizational culture and climate	√	√
Quality of implementation supports	√	√
Quality improvement of outcomes	√	√
Readiness	√	√
Service outcomes	√	√
Scale up	√	√
Spread	√	√
Sustainability	√	√
Transportability	√	
Variability of outcomes	√	√
Use of data	√	√

Research and evaluation design

Program evaluation uses both qualitative and quantitative research designs. But compared to other designs, randomized controlled trials (RCTs) have a very high degree of internal validity, which is crucial when it comes to assessing causation. While RCTs provide the greatest rigor for program evaluation, they also have drawbacks. Among them is the time it takes to reach conclusions about the effectiveness and impacts of an intervention. Furthermore, not all RCTs include detailed consideration of context or other factors affecting the quality of implementation of an intervention. Implementation science and improvement science argue for more practical and nimbler program development and for evaluation designs that can uncover the critical ingredients leading to successful implementation of early childhood interventions. Though some of these research design elements can be embedded in RCTs,[7] other innovative evaluation designs allow researchers, policymakers, and program designers to test innovations, identify important variability (Bryk, 2015), and get relatively quick answers to questions about what works for whom under what circumstances.

Mixed methods

Qualitative designs such as case studies are common when studying implementation of an intervention, yet many program evaluators and implementation scientists also use a combination of both qualitative and quantitative data sources, referred to as mixed methods, when studying implementation (Palinkas et al., 2011). For example, Nores and colleagues (2018) recently used a combination of qualitative and quantitative measures to track the early progress of an emergent, Reggio-inspired early childhood curriculum being implemented and scaled up in Columbia. Similarly, researchers interested in studying improvement also use qualitative or mixed methods. Indeed, Nores and colleagues state that the data they gathered and shared with program developers on processes around teacher training, observed quality of interactions in the classroom, and teacher perceptions of their work informed subsequent reforms in program policies and practices and changes to learning materials whose goal was to improve the quality of the curriculum and its implementation across the country.

[7] Examples include randomized cluster trials such as stepped wedge designs (Brown & Lilford, 2006; Gustafson et al., 2013; Hemming, Haines, Chilton, Girling, & Lilford, 2015) and pragmatic trials of all types. Pragmatic trials are controlled trials conducted in real-world, clinical settings (Peters et al., 2013; Roland & Torgerson, 1998). Multiphase Optimization Strategy (MOST) and Sequential Multiple Assignment Randomized Trial (SMART) are types of pragmatic trials that allow testing of implementation when one is initially developing an intervention (Collins, Murphy, Nair, & Strecher, 2005; Collins, Murphy, & Strecher, 2007). While pragmatic trial designs are relevant for a discussion of combining investigations of implementation and impact, a full consideration of all pragmatic design options is beyond the scope of this chapter.

Quasi-experimental designs

Quasi-experimental designs are often more practical and ecologically valid than RCTs for evaluating interventions in real-world settings. An evaluation design that is especially suited for implementation studies is the interrupted time-series experiment, which involves repeated assessments both before and after an intervention is implemented. This design is particularly helpful when evaluating the implementation of social policies (Biglan, Ary, & Wagenaar, 2000).

Other quasi-experimental designs that provide rigorous alternatives to a classic RCT include regression discontinuity and propensity score matching (Cappelleri & Trochim, 2015; Henry, Tolan, Gorman-Smith, & Schoeny, 2017). A regression discontinuity design assigns an intervention study's participants to treatment and control groups based on a pretreatment cutoff score (Cappelleri & Trochim, 2015). Distinct cutoff dates (such as that a child must reach age 5 by September 1 to be enrolled in kindergarten) or events (such as the mandated start date of a new state policy written into legislation) often serve as the point of discontinuity between those in and outside the treatment group. Propensity score matching, on the other hand, attempts to control for self-selection into an intervention by statistically matching participants and nonparticipants on a set of observed baseline characteristics that may represent confounding factors, such as level of educational attainment of parents or early childhood educators (Austin, 2011).

Innovative designs

Although many implementation and improvement studies to date are mainly descriptive in nature, several innovative evaluation designs permit the systematic examination of implementation within explanatory evaluation designs. These "blended" approaches allow the simultaneous examination of implementation processes and intervention outcomes (Granger et al., 2016; Granger & Shah, 2015; Nores et al., 2018; Peters et al., 2013; Pokorney, Taft, & Granger, 2015). An example of this blended approach is the effectiveness-implementation hybrid design, which seeks to explore the role of implementation in intervention impacts by embedding implementation questions (and thus measures of implementation outcomes) within effectiveness trials (Curran, Bauer, Mittman, Pyne, & Stetler, 2012; Granger et al., 2016; Peters et al., 2013). There are three types of hybrid designs. In the first, researchers modify an effectiveness trial to gather information on the intervention's delivery. In the second, they carry out simultaneous testing of an intervention and an implementation strategy. In the third, they test an implementation strategy while still gathering information on an intervention's effectiveness (Curran et al., 2012). Using a blended approach allows for simultaneous and systematic examination of both intervention and implementation effects and helps researchers avoid a Type III error—erroneously concluding that an intervention's core components were ineffective when the real reason benefits of the intervention were not detected was because the intervention was poorly implemented. Such hybrid

designs are not common in early care and education research and evaluation. However, implementation and impact evaluations have been combined for studying home visiting models' effectiveness for improving outcomes in early childhood.

Some of the newer research and evaluation designs are particularly suited to quality improvement and implementation evaluations because they emphasize and support innovation and adaptation, provide feedback in real time, and aim to produce context-specific understandings that inform ongoing innovation (Patton, 2009; Patton, 2010). For example, developmental evaluation, sometimes called *real-time evaluation*, *emergent evaluation, action evaluation*, or *adaptive evaluation*, is defined by Michael Patton (2009) as "asking evaluative questions and applying evaluation logic to support program, product, staff and/or organizational development." The evaluator, he notes, is "part of a team whose members collaborate to conceptualize, design and test new approaches in a long-term, ongoing process of continuous improvement, adaptation and intentional change," and his or her "primary function in the team is to elucidate team discussions with evaluative questions, data and logic, and facilitate data-based decision-making in the developmental process" (p. 41).

Developmental evaluation embeds evaluation activities within the implementation process; it is conducted for the benefit of the implementers rather than for compliance or quality assurance purposes. The evaluator is therefore seen as part of the implementation team, not an outside entity. Developmental evaluation is also meant to capture complex processes as they unfold in real time, rather than linear processes that are theoretically hypothesized and empirically tested (Patton, 2010). Developmental evaluations also often develop new measures to monitor progress toward emergent goals.

Rapid-cycle evaluation is a relatively new way of thinking about evaluation that aims to conduct evaluations of programs or policies quickly but still rigorously and at the same time provide information to implementers for continuous quality improvement purposes (Shrank, 2013). A key goal of rapid-cycle evaluation is to evaluate interventions regularly, starting soon after implementation, to allow for fast identification of opportunities for course correction and improvement. In this way, rapid-cycle evaluation follows a typical PDSA improvement cycle approach and is well suited to the task of assessing an intervention during the early implementation stage. With input from stakeholders, performance metrics are selected. These performance metrics are then collected, rapidly analyzed, and shared with implementers on a regular and iterative basis.[8]

[8] Although random assignment is not required for rapid-cycle evaluation, one could collect metrics on both a treatment and control group.

Precision research is another new evaluation framework that, like implementation science and improvement science, was first adopted in the health field (National Research Council, 2011). Precision medicine and precision public health both seek to predict and improve response to treatment by customizing health interventions for specific populations. Precision research has three main components: (1) partnerships that include many stakeholders who can design and test new strategies; (2) specificity in defining and measuring the intervention, in the desired outcomes, and in mediating pathways to those outcomes; and (3) efficient research designs such as rapid-cycle evaluation or usability testing.[9] Precision research breaks down a complex intervention into its component parts and systematically tests how individual elements (or combinations of elements) change outcomes for specific participants or under particular circumstances. Evaluators of early childhood interventions are beginning to use precision research to pinpoint which specific elements of a complex intervention are considered the essential "active ingredients" for achieving desired outcomes for specific populations or contexts (HARC Guidelines Task Team, 2018). Although precision research represents an innovation in program evaluation, it also has many elements in common with traditional program evaluation, as well as with implementation science and improvement science. For example, engaging multiple stakeholders in the testing of new strategies is similar to engaging multidisciplinary implementation teams in a quality improvement process, and the operationalization of the intervention and outcomes of interest along with efficient research designs corresponds to the focus on use of data and feedback loops in both implementation science and improvement science.

Summary: similarities and distinctions

There are many similarities among the aims, research questions, and research methods used across implementation science and improvement science. Program evaluators and researchers interested in implementation and/or quality improvement in early childhood settings are all interested in understanding the processes, contexts, and subgroup variations associated with the act of implementing an intervention aimed at achieving better outcomes for children and families. The implementation science and improvement science frameworks are largely compatible with one another, and the distinctions between them are few and subtle (see Table 1 and Table 2). It is perhaps easier to see the distinctions among different types of program evaluation and improvement science than between implementation science and improvement science.

[9] This information is summarized from the Home Visiting Applied Research Collaborative (https://www.hvresearch.org/precision-home-visiting/innovative-methods).

While many program evaluations focus on whether the intervention adheres to its design features, whether the service components were delivered and received, and whether intended outcomes of the intervention are achieved, improvement science is interested in identifying innovative ways to reach improved outcomes, in making adaptations to evidence-based practices to address the context, and in supporting individuals, teams, and organizations in the process of continuous improvement. In contrast to program evaluations that test the effectiveness of one or more well-defined intervention models at a time (i.e., effectiveness studies), improvement science posits that there are many pathways to the same goal of improved outcomes and that many small adjustments can be tested at the same time by different people within a team, organization, or collaborative. Although implementers should be guided by evidence-based practice, improvement science argues that they should also be free to experiment and innovate, provided that those innovations are compatible with research evidence. Importantly, researchers and practitioners with an improvement science perspective often note that not every change is an improvement. So improvement science is not about change for change's sake. Rather, its primary goals are creating a culture of learning and supporting organizational capacity and individual leadership for continuous improvement.

Because implementation science is the systematic study of how interventions and innovations are enacted in the real world, it is flexible enough—and comprehensive enough—to accommodate the study of fidelity to evidence-based practices (the hallmark of effectiveness and impact evaluations), as well as the study of innovative and adaptive quality improvement practices (the hallmark of improvement science). Implementation science has contributed to both program evaluation and improvement science by articulating a set of important concepts (e.g., implementation stages, implementation teams, use of data and feedback loops, implementation infrastructure, implementation outcomes) that collectively support both fidelity to an evidence-based practice and the appropriate adaptation of an evidence-based practice to new contexts or different populations. With the common aim of understanding the conditions under which improved outcomes are achieved and sustained, implementation science and improvement science are inherently compatible frameworks. Although their disciplinary origins, specific research questions, evaluation designs, and practical techniques may differ somewhat, they can mutually inform one another in practice, and both can contribute to program development and evaluation. Through some of the newer and innovative evaluation frameworks such as effectiveness-implementation hybrids, developmental evaluation, rapid-cycle evaluation, and precision research, it is becoming easier to meld implementation science, improvement science, and program evaluation.

APPLYING THE DIFFERENT APPROACHES TO EARLY CHILDHOOD INTERVENTIONS: THE EXAMPLE OF HOME VISITING

Now that we have explored the similarities and distinctions between implementation science and improvement science, I want to illustrate how they have been applied to the study of early childhood interventions using the example of home visiting models. Home visiting is a service delivery method rather than a specific intervention. Home visiting models aim to improve outcomes for pregnant women, newborns, and growing families by providing parent education, social support, and connections to community services. Many home visiting models have been developed, some targeting subpopulations such as first-time mothers, teen mothers, low-income families, or families with children with disabilities or chronic health conditions.

Traditional program evaluation

Home visiting models have been the subject of many traditional program evaluations over the years. For example, the Home Visiting Evidence of Effectiveness (HomVEE) project, supported by the U.S. Department of Health and Human Services, recently reviewed the research evidence for 20 home visiting models (Sama-Miller et al., 2018). HomVEE includes evidence of effectiveness from well-designed, well-executed RCTs and quasi-experimental designs. Most evaluations of home visiting models measure participant outcomes targeted by the interventions, such as parenting practices, family functioning, child health and development, maternal health and mental health, child abuse and neglect, or maternal life course outcomes such as deferral of subsequent births (Gomby, Culross, & Behrman, 1999; Sama-Miller et al., 2018). As models have matured, longer-term outcomes have been monitored, such as reductions in juvenile delinquency, family violence, crime, and family economic self-sufficiency (Sama-Miller et al., 2018).

Literature reviews in the journal *Future of Children* summarized findings from rigorous evaluations of home visiting models in 1993, 1999, and 2009 (Gomby et al., 1999; Howard & Brooks-Gunn, 2009; Olds & Kizman, 1993). The Winter 1993 issue reported mixed effects from over 30 home visiting models but concluded that this service delivery strategy was promising enough to warrant further expansion (Olds & Kizman, 1993). The Spring/Summer 1999 issue acknowledged the quick proliferation of home visiting programs in the few years since the last review and highlighted findings from six home visiting models that had been implemented nationally. Once again, findings for intended outcomes were mixed, and the magnitudes of positive impacts, when found, were modest (Gomby et al., 1999). Generally, significant findings were more prevalent for parent outcomes than for child outcomes. The Fall 2009 review focused on nine home visiting programs for infants and toddlers—six implemented in the U.S. and three implemented elsewhere—and also found mixed results (Howard & Brooks-Gunn, 2009). Furthermore, the 1999 review of six national home visiting models, noted variability in outcomes across subgroups of families both within and across home visiting models and across sites of implementation for the same home visiting model (Gomby et al., 1999). Similarly, the 2009 review identified variation in results by subgroup within models (Howard & Brooks-Gunn,

2009). The wide variability in results both across and within the models reinforced the idea that these home visiting models were unique in their structure and implementation even if their targeted outcomes were similar and therefore that findings could not be generalized across home visiting models, program sites, or populations (Gomby et al., 1999).

A meta-analysis of 60 home visiting programs conducted in 2004 similarly concluded that parents and children significantly benefited from home visiting programs compared to controls, but the effect sizes were small; also, no single program characteristic or design feature affected outcomes for children or parents consistently across the models (Sweet & Appelbaum, 2004). The most recent HomVEE review found variability in outcomes across the 20 home visiting models that met the inclusion criteria; however, two home visiting models, Healthy Families America and Nurse-Family Partnership (NFP), showed the most positive impacts across all eight outcome domains targeted by the models (Sama-Miller et al., 2018).[10]

In sum, although findings have been mixed, home visiting has had a greater impact on parent outcomes than on child outcomes, which is consistent with parents being the primary recipients of most home visiting content and contact.[11] When significant impacts on outcomes have been found for home visiting models, the effect sizes have been modest. This finding is understandable, too, when we consider the complex nature of the risk factors affecting the families most targeted by home visiting.

Despite the mixed results, home visiting continues to be viewed as a promising service delivery strategy that can yield benefits for low-income and at-risk families with young children. In fact, the evidence for home visiting as an effective early intervention method was considered strong enough that in 2010 the Patient Protection and Affordable Care Act stipulated the creation of the Maternal, Infant, and Early Childhood Home Visiting (MIECHV) program. MIECHV provides federal funding to states, territories, and tribal entities to implement evidence-based home visiting models that meet the needs of target populations within their areas.[12] Twenty-five percent of the total MIECHV funding is available for implementation and rigorous evaluation of "promising approaches" within home visiting that do not yet have a strong evidence base.

---

[10] Healthy Families America had one or more favorable impacts in each of the eight domains (considered either primary or secondary outcomes), and Nurse-Family Partnership had favorable impacts in seven out of eight outcome domains (considered either primary or secondary).

[11] Some have argued that combining home visiting models with other early intervention strategies directly targeting children may be especially beneficial (Gomby et al., 1999; Weiss, 1993).

[12] For more information, see https://mchb.hrsa.gov/maternal-child-health-initiatives/home-visiting-overview or https://mchb.hrsa.gov/sites/default/files/mchb/MaternalChildHealthInitiatives/HomeVisiting/pdf/programbrief.pdf.

Recently, a national evaluation of the MIECHV program, called Mother and Infant Home Visiting Program Evaluation (MIHOPE), released a report describing the services families received in the various MIECHV-funded home visiting programs and the characteristics of families, home visitors, local programs, other home visiting stakeholders, and communities associated with differences in the services families received (Duggan et al., 2018). A subsequent MIHOPE report shared findings about the families served and the implementation of the MIECHV-funded programs (Michalopoulos et al., 2019). In general, the MIECHV program has encouraged and supported the incorporation of implementation science and improvement science frameworks into traditional program evaluation at the national, state, and local levels through funding of the MIHOPE evaluation, state-led evaluations, the Home Visiting Applied Research Collaborative (HARC), and the Home Visiting Collaborative Improvement and Innovation Network (HV CoIIN). I describe some of this work in more detail in the sections that follow.

Implementation science

The primary recommendation of the 1999 *Future of Children* home visiting issue was that home visiting models should improve their implementation and quality of services; the second recommendation was that research should guide improvements in implementation and quality (Gomby et al., 1999). Since then, implementation of home visiting models has been studied for two more decades. Indeed, assessment of implementation fidelity and quality of home visiting program delivery are among the features included in the HomVEE project's recent review of home visiting models. Also, many of the state-led evaluations of MIECHV focus on implementation fidelity.

Much of the research on implementation of home visiting models has centered on *intervention fidelity*, including the number and frequency of home visits completed by home visitors compared to what the program model calls for, or the amount of intended content delivered—all representing different aspects of the intended dosage of home visiting services. Some evidence from meta-analyses suggests that as the number of hours of home visiting increases, the magnitude of the benefit increases relative to control families, and that a program with two or more visits per month has greater benefits than does less intensive home visiting programs (Nievar, Van Egeren, & Pollard, 2010; Sweet & Appelbaum, 2004). The most recent HomVEE review reported that all 20 home visiting models that met the inclusion criteria had minimum requirements for the frequency of home visits (Sama-Miller et al., 2018). In addition, 18 of the 20 models had specified content and activities for home visitors to use and had a system to monitor fidelity to content and activity.[13] However, another recent review of home visiting models noted that nine out of the 21 studies reviewed failed to indicate the duration of the home visits or how closely paraprofessional home visitors followed the program model (Peacock, Konrad, Watson, Nickel, & Muhajarine, 2013). Thus the level of information about intervention fidelity reported in the literature remains varied.

[13] The two home visiting models that lacked specified content were not the same two models that lacked a system to monitor fidelity to the content. See Table 4 in Sama-Miller et al. (2018) for further information.

The recent implementation study for MIHOPE provides more detailed information about implementation and the context for implementation than some previous studies (Michalopoulos et al., 2019). Families participating in MIHOPE received fewer home visits than expected by the evidence-based models, but they did receive a number of visits similar to what has been reported in previous studies of the models. Overall, 60% of participating families received at least half as many home visits as expected by their evidence-based models, a lower percentage than reported in previous studies (Michalopoulos et al., 2019).

Other research has examined *implementation fidelity*—that is, the evidence that implementation infrastructure and processes are in place and working well. Specifically, this research has examined the characteristics of home visitors and the training, ongoing support, and supervision necessary for effective implementation of a home visiting model (Tomlinson, Hunt, & Rotheram-Borus, 2018; Wasik, 1993). The recent HomVEE review noted that minimum education requirements for home visiting staff were specified by 17 of the 20 models reviewed; 18 models had minimum requirements for home visitor supervision; and all 20 models had preservice training requirements for home visitors (Sama-Miller et al., 2018). Selection, training, and ongoing supervision of staff are all part of the implementation infrastructure that supports implementation of an intervention such as home visiting. The implementation report for MIHOPE indicated that home visitors reported receiving more hours of training per month but fewer hours of individualized supervision per month than was expected by the evidence-based models (Michalopoulos et al., 2019). Inconsistent supervision and insufficient training are two of several "threats to implementation" that can affect delivery of an intervention model (Paulsell, Del Grosso, & Supplee, 2014).

Other aspects of this infrastructure include institutional policies and practices that facilitate the implementation of the intervention, partnerships that can help to sustain the intervention, data systems and use of data for ongoing monitoring and improvement, and the cultivation of leadership at all levels in support of the intervention (Aarons et al., 2011; Fixsen et al., 2005; Tomlinson et al., 2018). Less research has been published on these other aspects of implementation infrastructure, but they are just as vital to successful implementation as are the selection, training, and ongoing supervision that undergird staff competencies and intervention delivery.

One example that illustrates the important role of implementation infrastructure in supporting the implementation of an evidence-based home visiting model is the scaling up of the NFP home visitation model across the country that Dr. David Olds and his colleagues have undertaken (Hill & Olds, 2013). In the process of national scale-up, the program developers designed an initial set of implementation supports that focused on intervention fidelity and some aspects of implementation infrastructure such as staff competencies, financing, and data systems. Specifically, initial implementation supports included job descriptions for key staff; detailed guidelines and training for nurses and supervisors on the model's underlying philosophy and model elements; a startup guide for administrators to help plan for adequate and sustainable financing; and a data collection and reporting system to gather information on elements of program implementation (e.g., visit frequency, duration, and content), critical aspects of program

management (e.g., frequency of reflective supervision), and selected indicators of desired outcomes (e.g., tobacco and alcohol use during pregnancy, birthweight). However, as NFP began to be offered in new communities, the information provided by the data collection and reporting system quickly indicated that additional supports were necessary. Specifically, *organizational culture* needed to change: supervisors needed to recalibrate their expectations of a reasonable caseload for the nurse home visitors. Also, institutional *policies* (e.g., human resources policies and/or union rules) needed to be accommodated or amended to support the implementation of NFP in new communities.

Perhaps with new guidelines on reporting, more published journal articles will report on the implementation and improvement supports for early childhood interventions in the future.

In sum, Olds and colleagues recognized a need to address all aspects of implementation infrastructure to adequately support the successful implementation of the home visiting model in community-based settings at scale (Hill & Olds, 2013). They also understood the importance of *linked implementation teams* in the scaling process. In 2003, the developers established—with the support of several foundations—a national nonprofit to support national program implementation of NFP. As part of this system, regionally based NFP nurse consultants have access to feedback from the field through data system reports, and they address technical and adaptive challenges that arise in local implementing agencies as necessary (Hill & Olds, 2013). NFP is not the only home visiting model that has developed these additional implementation supports. Eighteen of the 20 models reviewed by the HomVEE project had established national headquarters to support local sites with implementing the model, and 15 had fidelity standards for local implementing agencies (Sama-Miller et al., 2018). However, few published reports of home visiting models provide detailed information about these implementation supports and how they function.[14] Perhaps with new guidelines on reporting, more published journal articles will report on the implementation and improvement supports for early childhood interventions in the future (Ogrinc, Davies, Goodman, Batalden, Davidoff, & Stevens, 2016; Yousafzai, Aboud, Nores, & Kaur, 2018).

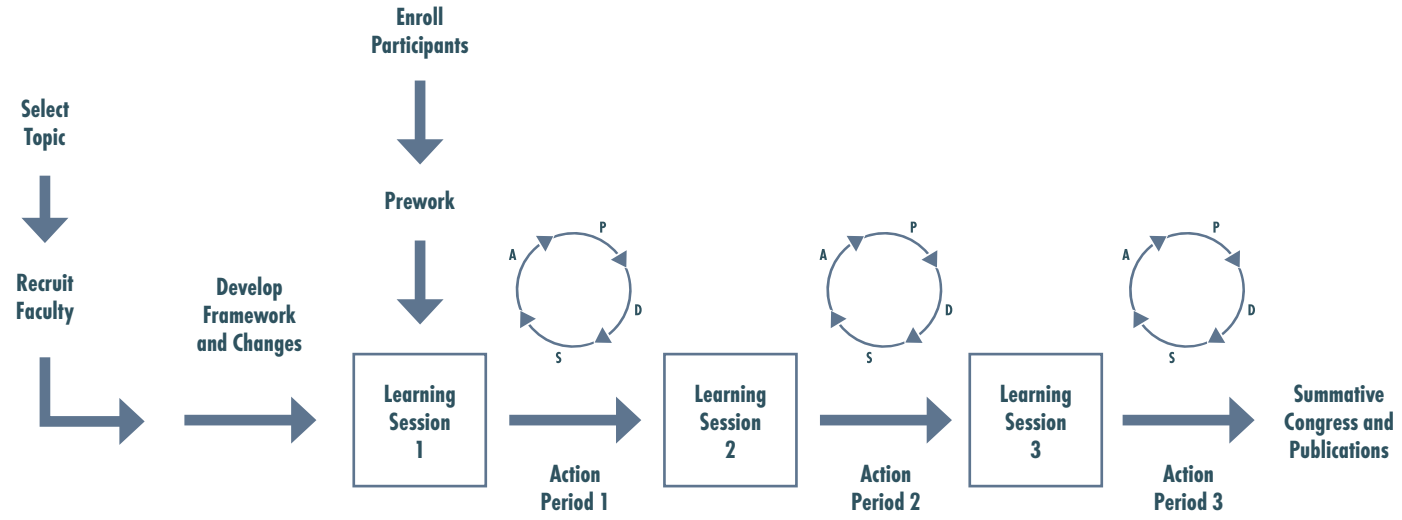
[14] As I have already noted, Hill and Olds (2013) thoughtfully reflected on the implementation infrastructure needed to scale NFP, but that was in a book chapter; such detail is not often found in journal articles. Olds (2006) also provides some information about implementation infrastructure, but not in as much detail.

Improvement science

The home visiting field has also embraced a focus on continuous quality improvement. In 2013, the HV CoIIN was established by the Health Resources and Services Administration (HRSA) to accelerate improvement among MIECHV grantees.

The CoIIN followed the BSC structure for continuous improvement (see Figure 2). As a first step in the development of the HV CoIIN, HRSA staff and others engaged in a topic selection process corresponding to the exploration stage of an implementation project. A group of subject matter experts convened in September 2013 to identify topics that would lead to improvement in home visiting outcomes. The goal was to identify topics that were aligned with MIECHV benchmarks, considered high priority by MIECHV grantees, and "ripe" for improvement (Mackrain & Cano, 2014). The experts identified three evidence-based topics (specifically, breastfeeding, developmental screening, and maternal depression) and the "innovative" topic of family engagement.[15]

Figure 2. Improvement science methodology.



Source: Reprinted from www.IHI.org with permission of the IHI, ©2018.

[15] They considered family engagement to be an innovative topic because it was deemed important but had less of an evidence base upon which grantees could draw for improvement.

The installation stage of implementation of the CoIIN began with assembling the HV CoIIN leadership team and faculty.[16] The HV CoIIN leadership team included a project officer from HRSA, a project director from a consulting organization (Education Development Center, Inc.), an improvement advisor with expertise in the BSC model, a faculty chair who would oversee the expert faculty, a CoIIN consultant, and an external evaluator (Mackain & Cano, 2014).

The HV CoIIN also had three faculty experts for breastfeeding, two for developmental surveillance and screening, four for maternal depression, and one for family engagement. Additional experts were brought in to facilitate the CoIIN process, including model developers, MIECHV technical assistance providers, evaluators and project officers, and state and local MIECHV implementers (Mackrain & Cano, 2014). The team proceeded with installation activities by developing change frameworks for each of the four topic areas and the enrollment of participants/teams in the CoIIN.[17]

In total, the HV CoIIN engaged multidisciplinary teams from 13 MIECHV awardees[18] and 36 local implementation agencies to work on improvements in child and family outcomes by testing evidence-based practices in breastfeeding, developmental screening and referrals, and maternal depression screening, and "promising practices" or innovations in family engagement (Mackrain & Cano, 2014).Each of the 13 multidisciplinary teams included federal, state, and local leaders and comprised, at a minimum, agency leads, day-to-day supervisors, MIECHV home visitors, and family recipients. Each team was asked to focus on one of the three evidence-based practice areas as well as family engagement during the CoIIN.

The "prework" activity of the HV CoIIN aimed to establish team identity, foster positive team dynamics and leadership among all team members, and introduce the change frameworks and quality improvement methods to the teams. The change framework for addressing maternal depression, for example, adopted five primary approaches for focusing improvement efforts: developing standardized and reliable processes for screening and response; creating a competent and skilled workforce to address maternal depression; establishing standardized and reliable processes for referral, treatment, and follow-up; encouraging active family involvement in maternal depression support; and developing a comprehensive data tracking system (HV-ImpACT webinar, 2017).

---

[16] The term "faculty" is part of the BSC framework and denotes subject matter experts who help guide collaborative teams in the use of evidence-based practices associated with a particular topic or activity. Both BSCs and CoIINs have higher-order implementation teams that help guide the collaborative teams and faculty. In this HV CoIIN, the implementation team was called the leadership team.

[17] Change frameworks are core elements of both CoIINs and BSCs. They delineate pathways for achieving improvements in topic-specific outcomes based on evidence (or best practice). Change frameworks identify the primary and secondary approaches for achieving the desired goals for a particular focal topic.

[18] The awardees included 10 states, two tribes, and one not-for-profit. Mackrain and Cano (2014) identify the number of MIECHV awardees for the first HV CoIIN as 13, but elsewhere it is recorded as 12 (see https://mchb.hrsa.gov/sites/default/files/mchb/MaternalChildHealthInitiatives/HomeVisiting/pdf/programbrief.pdf). It is possible that one team dropped out along the way.

Underneath each of these primary drivers lay a set of "secondary drivers," which were more specific, targeted activities related to the primary drivers. During the prework period, teams that had chosen maternal depression as their focus for the CoIIN could perform a self-assessment to help them determine which of the five primary drivers were already strengths and which could use improvement. This process helped the teams decide which of the primary and secondary drivers would be a starting point for their improvement work. The prework activity bridged exploration and installation stages, preparing the collaborative teams, faculty, and staff to begin active implementation of quality improvement activities.

The structured QI methodology of a BSC uses a series of learning sessions and action periods to accelerate improvements in the targeted topical areas (IHI, 2003). The HV CoIIN learning sessions were face-to-face meetings where faculty, staff, and collaborative teams shared information and ideas about evidence-based practices associated with the focal topics and further refined their understanding of quality improvement methods. For example, the teams learned about the Associates in Process Improvement's Model for Improvement (IHI, 2003), which uses PDSA cycles to answer three questions: What are we trying to accomplish? How will we know if a change is an improvement? What changes can we make that will result in improvement? Addressing these questions formed the basis of the work accomplished during the action periods. The collaborative teams identified what they hoped to accomplish by testing changes in practice related to breastfeeding, developmental screening, maternal depression, and/or family engagement. They also identified and refined performance metrics associated with these changes that were specific, measurable, achievable, relevant, and time-bound.[19] As a collaborative, the HV CoIIN agreed to the following performance metrics aligned to each of the four topic areas:

- Eighty-five percent of the women who screen positive for depression and access services will report a 25% reduction in symptoms in 12 weeks from first service contact.
- Increase by 25% from baseline the proportion of children with developmental or behavioral concerns receiving identified services in a timely manner.
- Increase by 20% from baseline the proportion of women exclusively breastfeeding at 3 and 6 months.
- Increase by 25% the average proportion of expected in-person contacts between home visitor and family that are completed.

[19] These characteristics go by the acronym S.M.A.R.T. and were first used in association with developing organizational goals and objectives (Doran, 1981). They should not be confused with the SMART design for intervention development discussed earlier (Collins et al., 2007).

During the action periods, collaborative teams tested their efforts in quality improvement in local settings using PDSA cycles to document their practice changes, reflect on their activities, and assess whether the changes in practice resulted in improvements in outcomes; they also gathered performance metrics associated with the target outcomes. Collaborative teams were supported in this process by the leadership team and faculty, who might initiate phone calls, send emails, conduct site visits, or host online discussion groups during action periods (see Figure 2).

Each member of a collaborative team used PDSAs and performance metrics during the action periods. For example, the state of New Jersey, one of the MIECHV grantees involved in the HV CoIIN, tested whether a phone call to prospective families from a home visitor would increase the number of families that enrolled in home visiting programs. The state agency collected and monitored data on enrollment rates at the state level while local home visiting programs collected performance indicators on enrollment rates in their programs (Supplee & Daily, 2018). Members of the New Jersey HV CoIIN shared data via an online dashboard that permitted individual programs to track and compare their performance over time and to see state-level aggregate data. This PDSA on the use of a phone call contributed to increased rates of enrollment in home visiting programs by almost 30% statewide (Supplee & Daily, 2018).

Three learning sessions and action periods occurred over 18 to 24 months. From an implementation stage-based perspective, the first learning session and action period would be considered part of early implementation, but subsequent learning sessions and action periods move collaborative teams toward full implementation of improvement practices and may even lead to spread and sustainability of such practices through changes in organizational culture (Bryk, 2015).

The HV CoIIN was active from September 2013 through August 2017. It demonstrated improvements in home visitors' knowledge and skills in the topical areas, as well as an increase in the use of data to achieve improvements in the targeted outcomes. However, it did not achieve the ambitious levels of performance hoped for across all performance metrics. For example, the rates of exclusive breastfeeding at 3 and 6 months rose only 3% instead of the hoped-for 20%. Specifically, exclusive breastfeeding at 3 months rose from 10% at baseline to 13.5% at the end of the CoIIN, and exclusive breastfeeding at 6 months rose from 5% at baseline to 8% at the end of the CoIIN (Arbour, Mackrain, Fitzgerald, & Atwood, 2018).

Nevertheless, the HV CoIIN was deemed successful in demonstrating that home visiting outcomes could be improved through this QI method, and many tools and resources were created through the HV CoIIN that could help spread and scale up improvement efforts among MIECHV grantees, potentially even those that had not participated in the CoIIN. As a result, a second, 4-year HV CoIIN (called HV CoIIN 2.0) was initiated in September 2017. HV CoIIN 2.0 will engage 25 state and territory MIECHV awardees and 250 local home visiting

agencies in quality improvement efforts around two topic areas that were addressed in the first CoIIN: (a) maternal depression screening, access to treatment, and symptom reduction, and (b) early detection of and linkage to services for developmental risk. In addition, the collaborative teams in HV CoIIN 2.0 will develop, test, and spread improvements in three new topical areas, the first of which is intimate partner violence.[20] Awardees will be selected in three waves. Each wave will last about 12 to 18 months and will once again use the BSC framework for quality improvement.

In sum, although improvements in performance metrics have been modest, positive qualitative outcomes associated with improvement science frameworks have led to additional investments in home visiting quality improvement collaboratives. Methods that focus on changing organizational climate to support continuous improvement seem promising compared to other quality improvement approaches that take a more individualized approach, such as one-on-one coaching. Early childhood researchers await with much interest and anticipation further evidence on the spread and sustainability of QI methods within organizations that participate in a BSC or CoIIN, as well as achievement of target performance metrics for the content addressed by these quality improvement models.

CONCLUSION

In this chapter, I argue that research methods relevant to the study of effective implementation and continuous quality improvement are compatible with methods used for early childhood program evaluation. Consequently, these frameworks can be easily combined in research and evaluation to support early childhood interventions. Furthermore, implementation science and improvement science frameworks, while distinct, are relatively similar and can inform one another.

To be most effective, implementation research methodology should be embedded within existing program and policy evaluation activities. For example, researchers can align their research and evaluation designs to the stage of implementation of an intervention or improvement model (Campbell et al., 2000; Permanency Innovations Initiative Training and Technical Assistance Project [PII-TTAP] & Permanency Innovations Initiative Evaluation Team [PII-ET], 2013). Taking an implementation perspective in program evaluation activities can provide a useful structure and may lead evaluators to look at processes and outcomes that otherwise might be left out of the equation. Focusing research attention on *who* is supporting the new practices and *how* they are providing that support (i.e., implementation teams and implementation infrastructure) is important because these aspects may be just as crucial to why an intervention achieved the outcomes it did as are components of the intervention and whether they were carried out with fidelity.

[20] For more information, see http://hv-coiin.edc.org/sites/hv-coiin.edc.org/files/HV%20CoIIN%20Information%20Resource%202017_0.pdf.

In short, implementation frameworks can help us understand why we get the results that we do for early childhood programs and policies. However, implementation frameworks should go beyond mere description and seek to explain the relationships among program or policy components and desired or expected outcomes as well. Some of the hybrid evaluation methodologies provide a promising approach to combining implementation science with effectiveness trials and impact evaluations.

A challenge that remains is embedding measures of implementation supports and implementation quality within program and policy evaluation models. Part of that challenge is the sheer number of variables that need to be considered in an expanded, more comprehensive program evaluation design that takes implementation into account (see Figure 1). Another challenge is the current dearth of rigorous measures of implementation. The development of valid and reliable measures that capture important elements of implementation and improvement is a keen pursuit for implementation researchers (Pokorney et al., 2015; Powell et al., 2017; Saldana, 2014; Shea, Jacobs, Esserman, Bruce, & Weiner, 2014). Future research in the early childhood field will hopefully benefit from new measures of implementation and improvement, as well as from related concepts such as readiness for change (Bumbarger, 2015; Halle, Partika, & Nagle, 2019). Furthermore, new reporting guidelines make it more likely that the implementation and improvement supports for early childhood interventions will be reported in sufficient detail in future journal articles (Ogrinc et al., 2016; Yousafzai et al., 2018).

As with implementation science, incorporating an improvement science approach within early childhood program development and evaluation potentially has great benefits. For example, usability testing is a research design that lets researchers use PDSA improvement cycles at the earliest stages of implementation and thereby improve and stabilize the essential functions and core components of a new intervention by testing just a few elements at a time (PII-TTAP & PII-ET, 2013). Rapid-cycle evaluation also uses PDSA cycles to provide frequent and ongoing feedback to program developers and evaluators.

Improvement science methods that emphasize interdisciplinary collaborative teams; that promote leadership at all levels of an organization; that support changes in organizational climate, and testing; and that document small practice changes collectively have been shown to lead to accelerated adoption of evidence-based practices. However, systematic reviews of quality improvement collaboratives note several limitations, including a lack of direct assessment of provider behavior and patient outcomes (there is, instead, heavy reliance on administrative data), and relatively few studies of cost effectiveness of the quality improvement models or sustainability of improvements over time (Nadeem et al. 2013; Schouten, Hulscher, van Everdingen, Huijsman, & Grol, 2008; Wells et al., 2017).

The promise of quality improvement methods such as BSC and CoIIN is beginning to be tested in home visiting (Arbour et al., 2018), publicly funded early education (Arbour et al., 2016), and community-based child care (Douglass, 2015; Hetzner et al., 2018). As the study of these methods continues in the early childhood field, we will need to consider whether collaborative improvement methods support more sustained and cost-effective improvements in outcomes compared to other quality improvement methods, such as coaching or professional learning communities.

While the investigation of the critical ingredients for improving the quality of early care and education and achieving the outcomes we want for young children is still a work in progress, we do know what some of those key ingredients are thanks to implementation science and improvement science. Rigorous program evaluation designs that permit comparisons of different types of program improvement methods—and that consider implementation processes, structures, and outcomes—will help the field further clarify what it takes to achieve improved outcomes for early childhood practitioners and settings, and for the children in their care.

ACKNOWLEDGMENTS

This chapter was supported by funding from the Foundation for Child Development. In addition to Sara Vecchiotti, Vice President for Research and Program Innovation at the Foundation for Child Development, I would like to thank Jennifer Agosti, Jan DeCoursey, Alex DeMand, Dean Fixsen, Kristen Keesee, Akiva Liberman, Rich Neimand, Lauren Supplee, and Jonathan D. Wallace for their review of and thoughtful comments on previous versions of this chapter. I would also like to thank Van-Kim Lin for her assistance with designing Figure 1.

References

Aarons, G. A., Hurlburt, M., & Horowitz, S. M. (2011). Advancing a conceptual model for evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(1), 4-23.

Agosti, J., Conradi, L., Halladay Goldman, J., & Langan, H. (2013, June). *Using trauma-informed child welfare practice to improve placement stability: Breakthrough Series Collaborative: Promising practices and lessons learned.* National Center for Child Traumatic Stress. https://www.nctsn.org/sites/default/files/resources//using_trauma_informed_child_welfare_practice_to_improve_placement_stability_practice_cards.pdf

Ammerman, R. T., Putnam, F. W., Margolis, P. A., & Van Ginkel, J. B. (2009). Quality improvement in child abuse prevention programs. In K. A. Dodge & D. L. Coleman (Eds.), *Preventing child maltreatment: Community approaches* (pp. 121-138). New York: Guilford.

Arbour, M., Mackrain, M., Fitzgerald, E., & Atwood, S. (2018). National quality improvement initiative in home visiting services improves breastfeeding initiation and duration. *Academic Pediatrics, 19*(2), 236-244.

Arbour, M., Yoshikawa, H., Atwood, S., Duran Mellado, F. R., Godoy Ossa, F., Trevino Villareal, E., & Snow, C. E. (2016, March 2-5). *Improving quality and child outcomes in early childhood education by redefining the role afforded to teachers in professional development: A continuous quality improvement learning collaborative among public preschools in Chile* [Paper presentation]. Society for Research on Education Effectiveness, Washington, DC. https://files.eric.ed.gov/fulltext/ED566990.pdf

Austin, P.C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3): 399-424.

Bauer, M. S., Damschroder, L., Hagedorn, H., Smith, J., & Kilbourne, A. M. (2015). An introduction to implementation science for the non-specialist. *BMC Psychology, 3*(1), 32.

Biglan, A., Ary, D., & Wagenaar, A. C. (2000). The value of interrupted time-series experiments for community intervention research. *Prevention Science, 1*(1), 31-49.

Boller, K., Sciarrino, C., & Waller, M. (2018, June 25-27). *The early learning social emotional engagement project: Collaborating for continuous learning and improvement* [Poster presentation]. National Research Conference on Early Childhood, Arlington, VA.

Brown, C. A., & Lilford, R. J. (2006). The stepped wedge design: A systematic review. *BMC Medical Research Methodology, 6*(54). https://bmcmedresmethodol.biomedcentral.com/track/pdf/10.1186/1471-2288-6-54

Bryk, A. S. (2015). Accelerating how we learn to improve. *Educational Researcher, 44*(9), 467-477.

Bumbarger, B. K. (2015). Readiness assessment to improve program implementation: Shifting the lens to optimizing intervention design. *Prevention Science, 16*(8), 1118–1122. https://link.springer.com/content/pdf/10.1007/s11121-015-0591-6.pdf

Cameron, E., & Green, M. (2009). *Making sense of change management: A complete guide to the models, tools, and techniques of organizational change* (2nd ed.). Philadelphia: Kogan Page.

Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D., & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. *BMJ, 321*(964), 694-696.

Cappelleri, J. C., & Trochim, W. M. (2015). Regression discontinuity design. In J. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd edition, Vol. 20, pp. 152-159). New York: Elsevier.

Century, J., & Cassata, A. (2016). Implementation research: Finding common ground on what, how, why, where, and who. *Review of Research in Education, 40*(1), 169-215.

Chambers, B. F., Glasgow, R. E., & Stange, K. C. (2013). The dynamic sustainability framework: Addressing the paradox of sustainment amid ongoing change. *Implementation Science, 8*(117). https://implementationscience.biomedcentral.com/track/pdf/10.1186/1748-5908-8-117

Collins, L. M., Murphy, S. A., Nair, V. N, & Strecher, V. (2005). A strategy for optimizing and evaluating behavioral interventions. *Annals of Behavioral Medicine, 30*(1), 65-73.

Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The Multiphase Optimization Strategy (MOST) and the Sequential Multiple Assignment Randomized Trial (SMART): New methods for more potent eHealth interventions. *American Journal of Prevention Medicine, 32*(5 Suppl), S112-S118.

Conradi, L., Agosti, J., Tullberg, E., Richardson, L., Langan, H., Ko, S., & Wilson, C. (2011). Promising practices and strategies for using trauma-informed child welfare practice to improve foster care placement stability: A Breakthrough Series Collaborative. *Child Welfare, 90*(6), 207-225.

Curran, G. M., Bauer, M., Mittman, B., Pyne, J. M., & Stetler, C. (2012). Effectiveness-implementation hybrid designs: Combining elements of clinical effectiveness and implementation research to enhance public health impact. *Medical Care, 50*(3), 217-226. http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC3731143&blobtype=pdf

Daily, S., Tout, K., Douglass, A., Miranda, B., Halle, T., Agosti, J., Partika, A., & Doyle, S. (2018). *Culture of Continuous Learning Project: A literature review of the Breakthrough Series Collaborative* (BSC). (OPRE Report 2018-28). Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science, 4*(50). https://implementationscience.biomedcentral.com/track/pdf/10.1186/1748-5908-4-50

Deming, W. E. (1986). *Out of the crisis.* Cambridge, MA: MIT Press.

Derrick-Mills, T., Sandstrom, H., Pettijohn, S., Fyffe, S., & Koulish, K. (2014). *Data use for continuous quality improvement: What the Head Start field can learn from other disciplines, A literature review and conceptual framework.* (OPRE Report 2014-77). US Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. https://www.urban.org/sites/default/files/publication/39061/2000079-data-use-for-continuous-quality-improvement.pdf

Doran, G. T. (1981). There's a S.M.A.R.T. way to write management's goals and objectives. *Management Review, 70*(11), 35-36.

Douglass, A. (2015, December 2-3). *Building capacity for organizational improvement in early care and education: The Breakthrough Series Collaborative approach* [Paper presentation]. Annual Meeting of the Child Care Policy Research Consortium, Washington, DC.

Duggan, A., Portilla, X. A., Filene, J. H., Crowne, S. S., Hill, C. J., Lee, H., & Knox, V. (2018). *Implementation of evidence-based early childhood home visiting: Results from the Mother and Infant Home Visiting Program Evaluation.* (OPRE Report 2018-76A). Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/sites/default/files/opre/mihope_implementation_report_2018_10_26_508b.pdf

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*(3-4), 327-350.

Ebert, L., Amaya-Jackson, L., Markiewicz, J. M., Kisiel, C., & Fairbank, J. A. (2012). Use of the Breakthrough Series Collaborative to support broad and sustained use of evidence-based trauma treatment for children in community practice settings. *Administration and Policy in Mental Health and Mental Health Services Research, 39*(3), 187-199.

Fixsen, D. L., & Blase, K. A. (2008). *The National Implementation Research Network.* Chapel Hill: University of North Carolina, Frank Porter Graham Child Development Institute.

Fixsen, D. L., Blase, K., Duda, M., Naoom, S., & Wallace, F. (2009). *Core implementation components, Research on Social Work Practice, 19*(5), 531-540.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication No. 231). University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network.

Franks, R. P., & Schroeder, J. (2013). Implementation science: What do we know and where do we go from here? In T. Halle, A. Metz, & I. Martinez-Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 5-19). Baltimore, MD: Brookes Publishing.

Glasgow, R. E. (2009). Critical measurement issues in translational research. *Research on Social Work Practice, 19*(5), 560-568.

Gomby, D. S., Culross, P. L., & Behrman, R. E. (1999). Home visiting: Recent program evaluations: Analysis and recommendations. *The Future of Children,* 9(1) 4-26.

Granger, B. B., Pokorney, S. D., & Taft, C. (2016). Analysis of the effectiveness of implementation. Pt. 3 of Blending quality improvement and research methods for implementation science. *AACN Advanced Critical Care, 27*(1), 103-110.

Granger, B. B., & Shah, B. R. (2015). Design and data collection. Pt. 1 of Blending quality improvement and research methods for implementation science. *AACN Advanced Critical Care, 26*(3), 268-274.

Grol, R., Baker, R., & Moss, F. (2002). Quality improvement research: Understanding the science of change in health care. *Quality Safety Health Care, 11*(2), 110-111. https://qualitysafety.bmj.com/content/qhc/11/2/110.full.pdf

Gustafson, D. H., Quanbeck, A. R., Robinson, J. M., Ford II, J. H., Pulvermacher, A., French, M. T., McConnell, K. J., . . . McCarty, D. (2013). Which elements of improvement collaboratives are most effective? A cluster-randomized trial. *Addiction, 108*(6), 1145-1157.

Haine-Schlagel, R., Brookman-Frazee, L., Janis, B., & Gordon, J. (2013). Evaluating a learning collaborative to implement evidence-informed engagement strategies in community-based services for young children. *Child & Youth Care Forum, 42*(5), 457-473.

Halle, T. G., Metz, A., & Martinez-Beck, I. (Eds.) (2013). *Applying implementation science in early childhood programs and systems*. Baltimore, MD: Brookes Publishing.

Halle, T. G., Partika, A., & Nagle, K. (2019). *Measuring readiness for change in early care and education.* (OPRE Report 2019-63). Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/sites/default/files/opre/cceepra_measuring_readiness_for_change_in_ece.pdf

Health Foundation (2011). *Evidence scan: Improvement science.* https://www.health.org.uk/sites/health/files/ImprovementScience.pdf

Health Information Technology Resource Center (2013). *Continuous quality improvement (CQI) strategies to optimize your practice: Primer.* https://www.healthit.gov/sites/default/files/tools/nlc_continuousqualityimprovementprimer.pdf

Hemming, K., Haines, T. P., Chilton, P. J., Girling, A. J., & Lilford, R. J. (2015). The stepped wedge cluster randomized trial: Rationale, design, analysis, and reporting. *BMJ, 350*(h391). https://www.bmj.com/content/bmj/350/bmj.h391.full.pdf

Henry, D., Tolan, P., Gorman-Smith, D., & Schoeny, M. (2017). Alternatives to randomized control trial designs for community-based prevention evaluation. *Prevention Science, 18*(6), 671-680.

Hetzner, N. P., Arbour, M. C., Douglass, A., Mackrain, M., & Agosti, J. (2018, June 25-27). *Continuous quality improvement in infant and early childhood programs: Approaches and critical implications for practice.* [Paper presentation]. National Research Conference on Early Childhood, Arlington, VA.

Hill, P., & Olds, D. (2013). Improving implementation of the Nurse-Family Partnership in the process of going to scale. In T. Halle, A. Metz, & I. Martinez-Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 193-207). Baltimore, MD: Brookes Publishing.

Howard, K. S., & Brooks-Gunn, J. (2009). The role of home-visiting programs in preventing child abuse and neglect. *The Future of Children, 19*(2), 119-146.

Hulleman, C. S., Rimm-Kaufman, S. E., & Abry, T. (2013). Innovative methodologies to explore implementation: Whole-part-whole–Construct validity, measurement, and analytical issues for intervention fidelity assessment in education research. In T. Halle, A. Metz, & I. Martinez-Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 65-93). Baltimore, MD: Brookes Publishing.

HV-ImpACT Webinar (2017). *Moving the bar for maternal depression: Systems strategies that work*. https://www.youtube.com/watch?v=e_G_jgzw9kc&feature=youtu.be

Institute for Healthcare Improvement (2003). *Innovation series 2003: The Breakthrough Series: IHI's Collaborative model for achieving breakthrough improvement* [White paper]. Institute for Healthcare Improvement.

Improvement Science Research Network (2010). *Research priorities*. http://www.isrn.net/research.

Langley, G. J., Moen, R. D., Nolan, K. M., Nolan, T. W., Norman, C. L., & Provost, L. P. (2009). *The improvement guide: A practical approach to enhancing organizational performance* (2nd ed.). San Francisco: Jossey-Bass.

Levinson W. A., & Rerick R. A. (2002). *Lean enterprise: A synergistic approach to minimizing waste*. Milwaukee, WI: ASQ Press.

Mackrain, M., & Cano, C. (2014, July). *Home visiting collaborative improvement and innovation network*. Maternal, Infant, and Early Childhood Home Visiting Technical Assistance Coordinating Center. http://hv-coiin.edc.org/sites/hv-coiin.edc.org/files/HV%20CoIIN%20Article%202014_0.pdf.

Martinez-Beck, I. (2013). Introduction: Where is the new frontier of implementation science in early care and education research and practice? In T. Halle, A. Metz, & I. Martinez-Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. xix-xxx). Baltimore, MD: Brookes Publishing.

McPherson, M. E., Gloor, P. A., & Smith, L. A. (2015). Using collaborative improvement and innovation networks to tackle complex population health problems. *JAMA Pediatrics, 169*(8), 709-710.

MEASURE Evaluation Working Group (2012). *Fundamentals of implementation research* (MS-12-55). MEASURE Evaluation. https://www.measureevaluation.org/resources/publications/ms-12-55

Metz, A., Halle, T. G., Bartley, L., & Blasberg, A. (2013). The key components of successful implementation. In T. Halle, A. Metz, & I. Martinez-Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 21-42). Baltimore, MD: Brookes Publishing.

Metz, A., Naoom, S. F., Halle, T., & Bartley, L. (2015). *An integrated stage-based framework for implementation of early childhood programs and systems* (OPRE Research Brief 2015-48). Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://nirn.fpg.unc.edu/sites/nirn.fpg.unc.edu/files/resources/OPRE-stage_based_framework_brief_508.pdf

Meyers, D. C., Durlak, J. A., & Wandersman, A. (2012). The quality implementation framework: A synthesis of critical steps to the implementation process. *American Journal of Community Psychology, 50*(3-4), 462-480.

Michalopoulos, C., Crowne, S. S., Portilla, X. A., Lee, H., Filene, H. J., Duggan, A., & Knox, V. (2019). *A summary of results from the MIHOPE and MIHOPE-Strong Start Studies of evidence-based home visiting* (OPRE Report 2019-09). Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.mdrc.org/sites/default/files/mihope_summary_brief_01_16_19_508.pdf

Nadeem, E., Olin, S. S., Hill, L. C., Hoagwood, K. E., & Horwitz, S. M. (2013). Understanding the components of quality improvement collaboratives: A systematic literature review. *The Milbank Quarterly, 91*(2), 354-394.

National Research Council (2011). *Toward precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease*. Washington, DC: National Academies Press.

Nievar, M. A., Van Egeren, L. A., & Pollard, S. (2010). A meta-analysis of home visiting programs: Moderators of improvements in maternal behaviors. *Infant Mental Health Journal, 31*(5), 499–520.

Nores, M., Figueras-Daniel, A., Lopez, M. A., & Bernal, R. (2018). Implementing aeioTU: Quality improvement alongside an efficacy study–learning while growing. A*nnals of the New York Academy of Sciences, 1419*(1), 201-217.

Ogrinc, G., Davies, L., Goodman, D., Batalden, P., Davidoff, F., & Stevens, D. (2016). SQUIRE 2.0 (Standards for Quality Improvement Reporting Excellence): Revised publication guidelines from a detailed consensus process. *BMJ Quality and Safety, 25*(12), 986-992.

Olds, D., Donelan-McCall, N., O'Brien, R., MacMillan, H., Jack, S., Jenkins, T., Dunlap, W. P., . . . Beeer, L. (2013). Improving the Nurse-Family Partnership in community practice. *Pediatrics, 132*(Supp. 2), S110-S117.

Olds, D. L., & Kitzman, H. (1993). Review of research on home visiting for pregnant women and parents of young children. *Future of Children, 3*(3), 53-92.

Øvretveit, J. (n.d.) *Interdisciplinary science of improvement and implementation.* https://globalimplementation.org/resources/interdisciplinary-science-of-improvement-and-implementation

Palinkas, L. A., Aarons, G. A., Horwitz, S., Chamberlain, P., Hurlburt, M., & Landsverk, J. (2011). Mixed method designs in implementation research. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(1), 44-53.

Patton, M. Q. (2009). *Developmental evaluation* [CES presentation]. https://evaluationcanada.ca/distribution/20090601_quinn_patton_michael_a.pdf

Patton, M. Q. (2010). Developmental evaluation: Applying complexity concepts to enhance innovation and use. New York: Guilford Press.

Paulsell, D., Del Grosso, P., & Supplee, L. (2014). Supporting replication and scale-up of evidence-based home visiting programs: Assessing the implementation knowledge base. *American Journal of Public Health, 104*(9), 1624-1632.

Peacock, S., Konrad, S., Watson, E., Nickel, D., & Muhajarine, N. (2013). Effectiveness of home visiting programs on child outcomes: A systematic review. *BMC Public Health, 13*(17). https://bmcpublichealth.biomedcentral.com/track/pdf/10.1186/1471-2458-13-17

Permanency Innovations Initiative Training and Technical Assistance Project & Permanency Innovations Initiative Evaluation Team (2013). *The PII approach: Building implementation and evaluation capacity in child welfare* (Rev. ed). Washington, DC: Department of Health and Human Services, Administration for Children and Families, Children's Bureau, and Office of Planning, Research, and Evaluation.

Peters, D. H., Adam, T., Alonge, O., Agyepong, I. A., & Tran, N. T. (2013). Implementation research: What it is and how to do it. *BMJ, 347*(f6753). https://www.bmj.com/content/bmj/347/bmj.f6753.full.pdf

Peters, D. H., Tran, N. T., & Adam, T. (2013). *Implementation research in health: A practical guide.* Geneva: Alliance for Health Policy and Systems Research, World Health Organization.

Peterson, S. (2013). Readiness to change: Effective implementation processes for meeting people where they are. In T. Halle, A. Metz, I. Martinez-Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 43-64). Baltimore, MD: Brooks Publishing.

Pokorney, S. D., Taft, C., & Granger, B. B. (2015). Analysis of the quality of implementation. Pt. 2 of Blending quality improvement and research methods for implementation science. *AACN Advanced Critical Care, 26*(4), 366-371.

Powell, B. J., Stanick, C. F., Halko, H. M., Dorsey, C. N., Weiner, B. J., Barwick, M. A., Damschroder, L. J., . . . Lewis, C. C. (2017). Toward criteria for pragmatic measurement in implementation research and practice: a stakeholder-driven approach using concept mapping. Implementation *Science, 12*(1), 118.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., . . . Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(2), 65-76.

Roland, M., & Torgerson, D. J. (1998). Understanding controlled trials: What are pragmatic trials? *BMJ, 316*(7127), 285.

Saldana, L. (2014). The stages of implementation completion for evidence-based practice: Protocol for a mixed methods study. *Implementation science, 9*(43). https://implementationscience.biomedcentral.com/track/pdf/10.1186/1748-5908-9-43

Sama-Miller, E., Akers, L. Mraz-Esposito, A., Zukiewicz, M., Avellar, S., Paulsell, D., and Del Grosso, P. (2018). *Home visiting evidence of effectiveness review: Executive summary* (OPRE Report 2018-113). Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/sites/default/files/opre/homvee_exec_summ_oct_2018_508.pdf

Schaack, D., Tarrant, K., Boller, K., & Tout, K. (2012). Quality rating and improvement systems: Frameworks for early care and education systems change. In S. L. Kagan & K. A. Kauerz (Eds.), *Early childhood systems*. New York: Teachers College Press.

Schouten, L. M. T., Hulscher, M. E. J. L., van Everdingen, J. J. E., Huijsman, R., & Grol, R. P. T. M. (2008). Evidence for the impact of quality improvement collaboratives: Systematic review. *BMJ 336*(7659), 1491-1494.

Scoville R., & Little K. (2014). *Comparing lean and quality improvement* [IHI white paper]. Institute for Healthcare Improvement.

Selk, S., Finnerty, P., Fitzgerald, E., Levesque, Z., & Taylor, J. (2015). Collaborative improvement and innovation network: Increasing capacity of data systems to help save babies' lives. *BMJ Quality & Safety, 24*(11) 738-739.

Shea, C. M., Jacobs, S. R., Esserman, D. A., Bruce, K., & Weiner, B. J. (2014). Organizational readiness for implementing change: A psychometric assessment of a new measure. Implementation *Science, 9*(7). https://implementationscience.biomedcentral.com/track/pdf/10.1186/1748-5908-9-7

Shojania, K. G., & Grimshaw, J. M. (2005). Evidence-based quality improvement: The state of the science. *Health Affairs, 24*(1), 138-150.

Shrank, M. (2013). The Center for Medicare and Medicaid Innovation's blueprint for rapid-cycle evaluation of new care and payment models. *Health Affairs, 32*(4), 807-812.

Supplee, L., & Daily, S. (2018). *Supporting continuous quality improvement in programs and services for young children*. https://www.childtrends.org/supporting-continuous-quality-improvement-in-programs-and-services-for-young-children.

Sweet, M. A., & Appelbaum, M. I. (2004). Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. *Child Development, 75*(5), 1435-1456.

Tansella, M., & Thornicroft, G. (2009). Implementation science: Understanding the translation of evidence into practice. *The British Journal of Psychiatry, 195*(4), 283-28.

Taylor, M. J., McNicholas, C., Nicolay, C., Darzi, A., Bel, D., & Reed, J. E. (2014). Systematic review of the application of the plan-do-study-act method to improve quality in healthcare. *BMJ Quality and Safety, 23*(4), 290-298. https://qualitysafety.bmj.com/content/qhc/23/4/290.full.pdf

Tomlinson, M., Hunt, X., & Rotheram-Borus, M. J. (2018). Diffusing and scaling evidence-based interventions: Eight lessons for early child development from the implementation of perinatal home visiting in South Africa. *Annals of the New York Academy of Sciences, 1419*(1), 218-229.

Tout, K., Epstein, D., Soli, M., & Lowe, C. (2015). A *blueprint for Early Care and Education Quality Improvement Initiatives. (Executive Summary)*. Bethesda, MD: Child Trends.

Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., . . . Saul, J. (2008). Bridging the gap between prevention research and practice: The interactive systems framework for dissemination and implementation. *American Journal of Community Psychology, 41*(3-4), 171-181.

Wasik, B. H. (1993). Staffing issues for home visiting programs. *Future of Children,* 3(3), 140-157.

Weiss, H. B. (1993). Home visits: Necessary but not sufficient. *Future of Children, 3*(3), 113-128.

Wells, S., Tamir, O., Gray, J., Naidoo, D., Bekhit, M., & Goldmann, D. (2017). Are quality improvement collaboratives effective? A systematic review. *BMJ Quality and Safety, 27*(3), 1-15.

Wesley, P. W., & Buysse, V. (Eds.) (2010). *The quest for quality: Promising innovations for early childhood programs.* Baltimore, MD: Brookes Publishing.

Young, B. (2017). *Continuous quality improvement in early childhood and school age programs: An update from the field.* Boston: BUILD Initiative.

Yousafzai, A. K., Aboud, F. E., Nores, M., & Kaur, R. (2018). Reporting guidelines for implementation research on nurturing care interventions designed to promote early childhood development. *Annals of the New York Academy of Sciences, 1419*(1), 26-37.

Yoshikawa, H., Wuermli, A. J., Raikes, A., Kim, S., & Kabay, S. B. (2018). Toward high-quality early childhood development programs and policies at national scale: Directions for research in global contexts. *Society for Research in Child Development Social Policy Report,* 31(1), 1-36.