



SECTION 3

**HOW DO WE GET SMARTER?
THE ROAD FORWARD**

IN SECTION 3:

Chapter 8: An Overview of Implementation Research and Frameworks in Early Care and Education Research.

By JoAnn Hsueh, Ph.D., MDRC, Tamara G. Halle, Ph.D., Child Trends, and Michelle Maier, Ph.D., MDRC

Chapter 9: Designing Implementation Research to Guide the Scale-Up of Effective Early Care and Education Across Settings.

By Michelle Maier, Ph.D., MDRC and JoAnn Hsueh, Ph.D., MDRC

Chapter 10: How Implementation Science and Improvement Science Can Work Together to Improve Early Care and Education.

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

Chapter 11: The Contributions of Qualitative Research to Understanding Implementation of Early Childhood Policies and Programs.

By Sharon Ryan, Ed.D., Rutgers, The State University of New Jersey

Chapter 12: Equity as a Perspective for Implementation Research in the Early Childhood Field.

By Milagros Nores, Ph.D., National Institute for Early Education Research



SECTION 3, CHAPTER 8

**AN OVERVIEW OF IMPLEMENTATION
RESEARCH AND FRAMEWORKS IN EARLY
CARE AND EDUCATION RESEARCH**

JoAnn Hsueh, Ph.D., MDRC

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

Michelle Maier, Ph.D., MDRC



Recent years have seen a healthy debate on the effectiveness of early care and education (ECE) programming, which includes home-based care providers, community-based child care centers, and publicly funded programs such as Head Start and prekindergarten. Some exemplary ECE programs have had substantial positive impacts on classroom quality and young children's learning and development at scale (e.g., Gormley, Phillips, & Gayer, 2008; Weiland & Yoshikawa, 2013). Some ECE programs also have the potential to narrow early achievement gaps experienced by children from low-income backgrounds (Gormley, Gayer, Phillips, & Dawson, 2005; Weiland & Yoshikawa, 2013), dual language learning children (Bloom & Weiland, 2015; Bumgarner & Brooks-Gunn, 2015), and children identified as having a racial or ethnic minority background (Currie & Thomas, 1999; Gormley et al., 2005; Weiland & Yoshikawa, 2013).

However, the literature also shows that ECE programs can vary in their overall effectiveness; they can be effective in one set of circumstances but not consistently so in others (Bloom & Weiland, 2015). ECE quality still varies considerably (Burchinal, Magnuson, Powell, & Hong, 2015), and not all efforts to enhance ECE quality ultimately improve children's outcomes, even when they show robust improvements on different dimensions of quality (Bryant et al., 2009; Pianta, 2013; Yoshikawa et al., 2015). Indeed, achievement gaps are substantial and persistent (Reardon & Portilla, 2016) and still emerge before children even step foot in kindergarten classrooms (Halle et al., 2009; von Hippel, Workman, & Downey, 2018).

In light of this promising but inconsistent evidence, increasing access to effective, high-quality ECE programming that reliably narrows achievement gaps is a pressing challenge. Important questions remain regarding how best to bring effective ECE programs to scale so that all children have access to high-quality learning experiences and so that investments in ECE programming ultimately close disparities in school readiness and achievement outcomes, as children move into and through formal schooling (Phillips et al., 2017).

To bring effective ECE programs to scale and ensure better outcomes for all children, an understanding of program implementation—that is, the process or specified set of steps by which a program is put into practice—as well as of variation in program implementation across contexts and populations is required. We also need to attend to internal and external factors that affect the quality of program implementation across contexts and at scale. Therefore, we see evidence that an ECE program is effective as necessary but insufficient to guide successful program scaling that benefits all children.

Research focused on implementation, particularly variation in implementation, can help address important knowledge gaps and issues in the ECE field.

Implementation-related activities include designing and articulating the critical components of a program model, identifying the supports needed to implement the model successfully, and understanding what drives variation in implementation across programs and participants and what it takes to transport an effective program to other contexts to meet the needs of diverse populations (Martinez-Beck, 2013). Research focused on implementation, particularly variation in implementation, can help address important knowledge gaps and issues in the ECE field regarding program evaluation, adaptation, expansion, and scale-up, including:

- **How to strengthen program effectiveness:** We need to know more about how effective ECE programs drive improvements in outcomes for children. Implementation research can help identify which program components are most critical for promoting which child outcomes—and for whom. These insights can be used to think about how programs can be optimized to produce reliable, positive impacts for young children and thereby narrow early disparities in achievement. Further, careful attention must be paid to ensuring that design and implementation of investments in ECE programming do not inadvertently reinforce or exacerbate existing inequities in our educational systems, which could have the effect of perpetuating or magnifying disparities in early achievement gaps (Nores, Ch. 12).
- **How to replicate results:** The processes and procedures that made a program successful in its initial context may not be the same for the program to be effective in another context (or for a different population). We need to understand more about how to transport and adapt promising ECE programs to new contexts while maintaining quality and effectiveness.
- **How to scale up:** Few effective ECE programs are operating on a large scale—that is, programming that reaches a broad population or is delivered across multiple contexts. As with replicability, the processes and procedures for taking an effective program and then adapting and expanding it to fit larger systems or to reach broader or more diverse populations are not well understood.
- **How to make programs sustainable:** The field often focuses on establishing systems and infrastructures to ensure the delivery of a program in line with its intended program model. Yet we still do not fully understand what it takes to ensure that a program is maintained in such a way as to allow it to continue to produce positive effects. We also need further study of where investments related to system infrastructure and program improvement should be focused to ensure that the program continues to narrow early disparities in achievement over time.

Implementation research is an important tool for illuminating what makes ECE programs, practices, and policies (collectively referred to as “programs” in this chapter) effective, what is needed to support program replication, expansion, and sustainability, and how to guide program improvement to help ensure that ECE programs reach their potential for narrowing achievement gaps. This chapter lays the groundwork for ensuing chapters and outlines principles and frameworks from implementation science that undergird implementation research of ECE programming.

WORKING DEFINITIONS OF IMPLEMENTATION SCIENCE AND IMPLEMENTATION RESEARCH

Implementation science is the set of frameworks and principles that explains the processes by which programs, policies, and individual practices are enacted in real-world settings (e.g., Century & Cassata, 2016; Damschroder et al., 2009; Peters, Adam, Alonge, Agyepong, & Tran, 2013). Implementation research encompasses the application of implementation science frameworks and principles to systematic inquiry into the act of carrying out a program, as well as systematic inquiry into how a program is received and experienced in real-world settings and situations. In its most basic form, implementation research and analysis aim to illuminate *what* is happening, *how* it is happening, *who* is making it happen, *why* a program achieves the outcomes that it does, and *for whom* it works best. Implementation research can take a vertical perspective, looking at how processes across different levels of the supporting system can work in synergistic or countervailing ways to support a program’s implementation, or it can take a horizontal perspective, examining how implementation unfolds across a range of different settings, contexts, and populations (Ryan, Ch. 11; Vavrus & Bartlett, 2006). Accordingly, implementation research can cover a wide range of topics, thereby providing an understanding of ECE programming at different stages of implementation and program development.

In its most basic form, implementation research and analysis aim to illuminate what is happening, how it is happening, who is making it happen, why a program achieves the outcomes that it does, and for whom it works best.

ADOPTING AN INWARD AND OUTWARD FOCUS ON IMPLEMENTATION

Implementation frameworks underscore where research can focus and, in turn, generate hypotheses and research questions. A growing set of implementation frameworks have been applied to ECE; one kind focuses inward on program components and structure, and another focuses outward on the contexts and larger infrastructure that support successful implementation of programs and systems. An inward focus articulates key aspects of implementation, such as core program components, implementation drivers, implementation processes, or different stages of implementation and program development (e.g., The National Implementation Research Network). An outward focus conceptualizes which features of larger systems may help expand programs that were previously evaluated on a small scale and considers how such programs may be scaled up with fidelity (Fixsen & Blase, 2008; Supplee & Metz, 2015). Theoretical models of implementation emphasize the interdependency of factors across levels of analysis, that is, at the level of the individual, organization, and larger systems (Aarons, Hurlburt, & Horwitz, 2011; Domitrovich et al., 2008; Fixsen, Blase, Metz, & Van Dyke, 2013). Given this interdependence, implementation researchers differ in their perspectives of what constitutes an inward or an outward focus. Indeed, these distinctions can shift with a researcher's focus of inquiry. For the purposes of this chapter, implementation research that focuses inward addresses a program's theory of change or implementation processes, while implementation research that focuses outward is oriented to the larger context and infrastructure supports that surround a program. These foci highlight potential sources of variation that may account for the effectiveness (or lack thereof) of ECE programs, as well as for how such programs may have varying effects in different contexts and for children with different backgrounds.

► Inward focus

Taking an inward focus means conducting a systematic inquiry into the program itself. This inquiry begins by articulating the underlying logic model and theory of change delineating the mechanisms by which the program yields improvements in short- and longer-term outcomes for children. The assumption here is that the program under study has been or can be defined so that its components, staffing, and features are recognizable (and replicable). When such a program is studied, the underlying logic model of the program then begins with the well-articulated, measurable, and recognizable program components and staffing, that is, the program that was planned. From there, the implementation of program components—the program that is offered to participants—can be distinguished from the program components received (or taken up) by participants. Another component of the inward focus on implementation is the role of the implementers, that is, those who carry out the program components within the program itself. Implementers can be a team of individuals or just one person, depending on the program parameters and structure.

Intervention fidelity is the process by which the program as offered and as received is evaluated in comparison to the program as planned (Dunst et al., 2008; Fixsen, Naoom, Blase, Friedman, & Wallace, 2005). It is important to note that intervention fidelity is a multidimensional construct that includes assessment of dosage, adherence, and quality, among others, to varying degrees in the literature (Dane & Schneider, 1998; Durlak & Dupre, 2008).

A focus on intervention fidelity provides a framework for inward examination of a program's theory of change or implementation processes. Such a focus is important because evidence indicates that variation in intervention fidelity influences outcomes (Durlak & DuPre, 2008; Wilson, Lipsey, & Derzon, 2003) and may lead to variation in program effectiveness. Further, in the context of program evaluation, understanding intervention fidelity is essential to interpreting outcomes. Without being able to assess implementation processes and fidelity, it is difficult to account for null or negative program effects. This is because it is not possible to parse whether null effects may be attributed to a lack of program strength (that is, poor intervention fidelity leading to no impacts) or to a poor program theory (that is, strong intervention fidelity but no impacts) (Dusenbury, Brannigan, Falco, & Hansen, 2003). In addition, assessing fidelity can help explain the why behind the causal relationships demonstrated through program impacts, as well as suggesting the effects that modifications to implementation processes and barriers to intervention fidelity may have on outcomes (Munter, Wilhelm, Cobb, & Cordray, 2014).

► **Outward focus**

Several conceptual frameworks guiding implementation research draw attention outward and focus on the broader organizational infrastructure, system, and/or contexts that influence implementation of a program model. Collectively, these systems and contexts have the potential to create a hospitable environment that can facilitate a program being carried out as expected (Fixsen et. al., 2005, Fixsen, Blase, Naoom, & Wallace, 2009; Metz, Bartley, Ball, Wilson, Naoom, & Redmond, 2015).

Elements of an outward focus on implementation include the implementation infrastructure (the tools, resources, and supports put in place to deliver the program model and underlying components), the implementation teams (organizations, providers, and individuals that help make successful delivery of the program model possible by supporting the implementers), and the characteristics of participants and contexts. These core elements can be conceptualized as proximal or distal contextual influences that interact dynamically with one another and also with the program itself.

These outward elements can operate in synergistic or countervailing ways to achieve desired outputs (delivery and receipt of program services by participants in line with the program model as planned) and, in turn, short- and longer-term outcomes for children. These contextual, organizational, and systems-level elements that support implementation represent an important source of variation that should be considered when evaluating the effectiveness of a program.

In order to deliver a program effectively as planned, with the ultimate goal of achieving outcomes for children that the early childhood program is designed to address, a strong infrastructure must be put in place to support the individuals who will carry out the underlying components of the program with fidelity. The implementation infrastructure includes organizational resources (both financial and in kind) that will provide for any materials and staff training required to implement the program, organizational policies and procedures that support rather than

work against the effective implementation of the program, external partnerships that will support the program and the organization in which it is embedded, strong leadership at all levels of the organization that will champion the program, and well-trained staff to carry out the program (Metz, Halle, Bartley, & Blasberg, 2013). This implementation infrastructure is sometimes categorized into three interrelated elements (Fixsen et al., 2005; Fixsen & Blase, 2008):

In order to deliver a program effectively as planned, with the ultimate goal of achieving outcomes for children that the early childhood program is designed to address, a strong infrastructure must be put in place to support the individuals who will carry out the underlying components of the program with fidelity.

- *competency drivers*, which refer to organizational processes that directly support the development and maintenance of the competency of frontline staff (including the selection, training, and continuous oversight and assessment of staff who are implementing the program), enabling them to carry out the program as planned,
- *organizational drivers*, which refer to the operating organization’s infrastructure and institutional capacity to support staff in implementing programs with fidelity (including policies and practices such as coaching) by using data and technology to monitor the progress of implementation of the program’s components, funding and other resources, and external partnerships that can provide additional resources for the effort, and
- *leadership drivers*, which refer to the individuals who are charged with supporting program implementation (but can—and should—also include those who are charged with direct program implementation) who can address both technical and behavioral/adaptive challenges to implementation.

These elements of the implementation infrastructure are hypothesized to be integrated and compensatory, meaning that if there is weakness in one area (e.g., you have limited control over the staff you can select to carry out the new practice), it may be possible to strengthen another area (e.g., you can offer additional training or coaching to existing staff and institute new organizational policies to support staff in the new practice) without compromising the overall supporting implementation infrastructure.

The people who support those who are implementing a new policy or practice are considered to be members of implementation teams (Halle et al., 2015). They are actors or teams who vary in terms of their power, influence, and proximity to the implementation of key program components. Examples include politicians and elected officials, who are generally further from the program's on-the-ground implementation; key personnel (such as program administrators and early adopters among program staff); and key stakeholders, such as program developers, who see themselves as authorizing an initiative or being responsible for the success of the initiative and take an active role in providing support for delivering the program components. Those who support implementation teams may do so in a variety of ways, such as by training individuals who are tasked with carrying out the new practice, monitoring success in carrying out the new practice, and/or providing feedback to practitioners to continually improve the new practice. Or they may be involved in funding the initiative, setting up and implementing supporting policies and practices within the organization, or creating alliances with partner organizations. With complex initiatives, multiple implementation teams supporting implementation at different levels of a program or system may be involved to provide the necessary leadership support.

An outward focus to implementation also considers the effect of the characteristics of the participants implementing and receiving the program, as well as the larger context in which the program is being implemented, on the success of program implementation. The composition of the participants and the relevant contextual characteristics may vary with regard to geography, reach, and scale of a program.

Whom the program intends to reach, as well as the population that is ultimately recruited, enrolled, and served, can vary. These are important considerations because a program that is effective for one group may not be effective for another. For example, the program may make certain assumptions about the risks, readiness, and capacities of intended participants. If the participants enrolled in the program do not bear out those assumptions, the model as it unfolds in real-world settings may need to be modified.

Similarly, a program that is effective in one set of contextual circumstances may not be effective in other circumstances, necessitating adaptation in key program components or adjustments in the implementation infrastructure. Contextual characteristics—such as political, economic, and social realities and constraints—can inform and shape implementation processes and infrastructures. Examination of context can bring to light other ways the program under study might serve those who are offered and receive it; a program's uptake in a community, and thus its ultimate "reach" and effectiveness, can vary depending on what other experiences are available to potential program participants in the area. In sum, characteristics of both program participants and settings offer critical insights into understanding a program's effectiveness.

While intervention fidelity is an important consideration for an inward focus to implementation, implementation fidelity is important for an outward focus. Implementation fidelity refers to the degree to which the implementation infrastructure and the supports encompassed therein—such as professional development, technical assistance, and other administrative assistance—are provided in a way that is consistent with what was planned. In some instances, resources and delivery of professional development supports may be distributed unevenly across a broad system of ECE programming. For example, the kinds of preparation and qualifications deemed necessary for and received by ECE teachers varies widely across ECE settings (see Pianta and Hamre, Ch. 5).

INTERSECTION WITH STAGE-BASED FRAMEWORKS OF IMPLEMENTATION AND PROGRAM DEVELOPMENT

Areas of exploration and inquiry related to an inward or outward focus on implementation can help specify the who, what, and how of program implementation as well as why, for whom, and under what circumstances a program is effective when delivered in real-world settings (e.g., Fixsen et al., 2005). These insights are critical when programs evolve and progress over time. Yet all too often, systematic inquiry of implementation, particularly from an outward perspective, becomes the focus of research only in later stages of implementation and program development. Such insights, however, can be instrumental even in early stages of implementation and program development; they can help the field understand how programs can ensure the effectiveness and quality of ECE programming for all children by strengthening and adapting themselves. Illuminating the extent to which there is or is not cohesion and alignment across these drivers of implementation can improve the development, scaling, and sustainability of ECE programming with diverse providers and staff and diverse groups of children and families and thereby help reduce disparities in early achievement gaps.

Advancing these efforts requires tying together systematic, stage-based inquiry of implementation and program development. Two often-referenced stage-based frameworks are especially relevant here.

► Stages of implementation

Several implementation frameworks identify multiple stages in the implementation process (Aarons, Hurlburt, & Horowitz, 2011; Meyers, Durlak, & Wandersman, 2012). The National Implementation Research Network, for instance, identifies four implementation stages: exploration, installation, initial implementation, and full implementation (Bertram, Blase, & Fixsen, 2015).

During *exploration*, stakeholders are assessing their needs and identifying what will best fit those needs in terms of adopting new programs, policies, or practices. They are also examining the feasibility of taking on a new practice, program, or policy, including assessing buy-in by all those affected by such a decision. During *installation*, the new program is not yet being delivered, but stakeholders are busy making sure that they have the technical, financial,

and human resources to carry it out. This may involve hiring and training new staff or training existing staff (i.e., addressing staff competencies) or making structural and instrumental changes organizationally (i.e., addressing organizational infrastructure) that enable stakeholders to carry out the new program. *Initial implementation* signals the start of service delivery. During this stage, data are regularly gathered and used to assess how well things are going and to make adjustments, as necessary, with the goal of continuously improving implementation. Rapid-cycle problem solving becomes prominent during this stage and continues even when full implementation is achieved. *Full implementation* is characterized by skillful implementation of the new program, with the necessary skilled practitioners, organizational infrastructure, and leadership in place to support its continued reliable use and sustainability.¹ While these stages are presented here in a sequential, linear order, there is consensus in the field of implementation science that the stages are recursive (Saldana, 2014), and that achieving full implementation of a well-defined, evidence-based program can take between two and four years (Bierman et al., 2002; Fixsen, Blase, Timbers, & Wolf, 2001).

► Stages of program development

Those involved in program development also use a stage-based framework to describe the process. This framework begins at an *early* or *developing* stage (before scale-up) with a program model that is new or recently developed. The program is often piloted on a smaller scale or in a relatively controlled setting (for example, under the direct supervision of its developers and with eager volunteer participants) with the aim of clarifying and, if necessary, refining the program goals, target population, and key activities and components as they are being implemented.

As a program matures, it may move through the stages of *promising* to *effective*, if early efficacy trials establish evidence of effectiveness when the program is delivered on a relatively small scale. At this stage, efforts typically focus on replicating prior results and/or expanding the program, that is, scaling up in a limited way, so that it can be tested in more diverse populations and contexts; this is called “horizontal scaling” (Dunst, Bruder, Trivette, & Hamby, 2006; Hartmann & Linn, 2008). Goals for program development may thus move on to tasks aimed at understanding whether, when, how, and for whom—meaning under what conditions, across what contexts, and with what populations—the program can be expanded or successfully replicated, while seeking to further test the program’s effectiveness.

¹ Some implementation science researchers identify *Sustainability* as a distinct, fifth stage or “phase” of implementation (Saldana, 2014). Similarly, a well-established implementation framework in health science research, RE-AIM, identifies *Maintenance* as the final component of implementation (Damschroder et al., 2009).

As the program matures further, it often moves to a *scaling* stage of program development, whereby it is scaled more extensively with the explicit goal of building the level of effectiveness evidence for institutionalizing the program into an existing system to ensure longer-term sustainability; this is called “vertical scaling” (Dunst et al., 2006; Hartmann & Linn, 2008).

At different stages of implementation and program development, insights gained from implementation research can help the field understand how programs can continue to strengthen and evolve, helping them ensure that effective, high-quality ECE programming is being delivered across localities and on a broad scale in an effort to narrow achievement gaps. For example, during an initial implementation stage, the goal is to monitor and continuously improve implementation and refine and strengthen program design. At this stage of implementation, implementation research focusing inward may gather data to assess how well the program is being implemented and how the experiences of children with different backgrounds or experiences might vary, which can then be used to identify areas in which implementation processes and/or the program model can be adjusted, as necessary. Implementation research that focuses outward at this implementation stage, in contrast, may gather data to assess how well the infrastructure system and implementation teams are supporting implementation and how these experiences might be influenced by the characteristics of staff, information that can then be used to make adjustments to those supports, as necessary. In turn, this information could be used to ensure that the program delivery does not inadvertently reinforce processes that contribute to disparities in early achievement skills.

Similarly, in early or developing stages, key aims are to refine the program goals, model, and target population. Feasibility studies, demonstrations, pilot assessments, and early efficacy tests are aligned with these goals and may help challenge assumptions about elements of the program that are essential as designed or encourage exploration of alternative approaches and strategies that could strengthen the program’s overall effectiveness. Implementation research in these earlier stages of program development may thus focus inward to assess intervention fidelity and explore how it may change with adjustments to the program model or characteristics of the population being served. Meanwhile, implementation research with an outward focus may begin to describe the intersectionality of setting characteristics, the implementation teams, and children being served with a goal of improving how resources or supports can be allocated and tailored to ensure high-quality learning experiences for all children as the program moves into different stages of development.

Later stages of program development may use similar types of tests (e.g., efficacy or effectiveness studies), but they have a different goal in mind. For example, at a scaling stage of program development, the foci of research may turn outward toward testing and mapping multiple levels of system, infrastructure, and institutional supports and describing tensions and alignment of these components that support ECE programming. Research may also focus on the variation in implementation and on illuminating variation in program impacts across contexts, populations, and conditions using a variety of qualitative and quantitative methodological approaches. Thus, blending implementation

research from both inward and outward perspectives, while situating a program along different stages of implementation and program development, can help to identify sets of research questions and evidence-building research activities that can be used to build ECE programming on a large scale that moves toward the ultimate aim of reducing disparities in early academic achievement.

CONCLUSION

The implementation frameworks we've presented illustrate where implementation research in ECE can continue to push forward in the coming years. By taking both an inward and outward perspective on implementation processes, research can point out how diversity in context, populations, resources, and systems intersect to affect the quality of ECE programming and in turn can broaden our knowledge of the influences that shape the lives and trajectories of children and that contribute to noted disparities in achievement as children progress through schooling. Research to date has provided some insights into the sources of variation in the effectiveness of different ECE programs. But many of the contextual influences that may lead to variation in an ECE program's effectiveness, particularly when delivered on a large scale, remain to be studied.

Implementation frameworks serve as organizing tools that help highlight underexplored areas and point to ways to improve ECE program effectiveness for narrowing achievement gaps. These frameworks suggest the need for more systematic collection of data early on about factors that constitute the supports for implementation and for a broadening of the conceptualization of measures and research designs that aim to address questions at different stages of implementation and program development.

Further, stage-based approaches to implementation research can be incorporated into the development, implementation, and scaling of effective early childhood programs, practices, and policies, with the research feeding back into ongoing improvement, sustainability, and scaling activities. By embedding the study of ECE programs within these frameworks, we can begin to broaden our knowledge of the influences that shape the lives and trajectories of young children, particularly those from low-income and racial, ethnic, and immigrant minority backgrounds.

By embedding the study of ECE programs within these frameworks, we can begin to broaden our knowledge of the influences that shape the lives and trajectories of young children, particularly those from low-income and racial, ethnic, and immigrant minority backgrounds.

Succeeding chapters build on the implementation frameworks introduced here and extend the conversation beyond the immediate impacts of ECE programming to more in-depth discussions and illustrations of how implementation research can be applied in innovative ways to guide and strengthen ECE programming and practices for all children.

“Designing Implementation Research to Guide the Scale-Up of Effective Early Care and Education Across Settings,” by Michelle Maier and JoAnn Hsueh, describes a framework that can help guide the empirical study of program implementation within an evidence-building context and discusses potential methodological and measurement considerations researchers should bear in mind when adopting an inward and outward focus to implementation research as a means of understanding variation in the impacts of ECE programming across diverse populations, contexts, and conditions.

In her chapter, “How Implementation Science and Improvement Science Can Work Together to Improve Early Care and Education,” Tamara G. Halle outlines the similarities and distinctions between implementation science and improvement science. The chapter provides concrete examples of these approaches as they have been applied to the study of home visiting models as a form of early childhood intervention aimed at improving outcomes for children and families. It concludes by considering how integrating implementation science, improvement science, and traditional program evaluation can further support the effectiveness and sustainability of early childhood interventions, especially those targeted to ECE settings.

Sharon Ryan’s chapter, “The Contributions of Qualitative Research to Understanding Implementation of Early Childhood Policies and Programs,” discusses qualitative methods that researchers can draw on to understand how processes of implementation are constructed and adapted. It underscores the value of moving beyond children’s immediate experiences in the classrooms, to take into account the perspectives of local actors, conditions, and contexts, and to begin to theorize how ECE policies, systems, and programs can be improved to address the needs of children with diverse backgrounds.

Milagros Nores’s chapter, “Equity as a Perspective for Implementation Research in the Early Childhood Field,” underscores that researchers must tackle biases and cultural limitations introduced by their own research methods; doing so will enable them to appropriately and fully understand how programs are operated and implemented across settings, contexts, and populations with diverse histories and backgrounds. This information can be used to assess the degree to which ECE programming meets equity goals of reducing inequity in young children’s learning opportunities and experiences.

References

- Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(1), 4-23.
- Bertram, R. M., Blase, K. A., & Fixsen, D. L. (2015). Improving programs and outcomes: Implementation frameworks and organization change. *Research on Social Work Practice*, 25(4), 477-487.
- Bierman, K. L., Coie, J. D., Dodge, K. A., Greenberg, M. T., Lochman, J. E., McMahon, R. J., & Pinderhughes, E. (2002). The implementation of the Fast Track program: An example of a large-scale prevention science efficacy trial. *Journal of Abnormal Child Psychology*, 30(1), 1-17.
- Bloom, H. S., & Weiland, C. (2015). *Quantifying variation in Head Start effects on young children's cognitive and social-emotional skills using data from the National Head Start Impact Study*. New York, NY: MDRC.
- Bryant, D. M., Wesley, P. W., Burchinal, M., Sideris, J., Taylor, K., Fenson, C., & Iruka, I. U. (2009). *The QUINCE-PFI Study: An evaluation of a promising model for child care provider training: Final report*. Chapel Hill, NC: Frank Porter Graham Child Development Institute, University of North Carolina.
- Bumgarner, E., & Brooks-Gunn, J. (2015). The association between early care arrangements, quality, and emergent bilingual Latino American children's math and literacy skills in English. *Early Childhood Research Quarterly*, 30(1), 32-44.
- Burchinal, M., Magnuson, K., Powell, D., & Hong, S. S. (2015). Early childcare and education. *Handbook of Child Psychology and Developmental Science*, 4(6), 1-45.
- 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.
- Currie, J., & Thomas, D. (1999). *Early test scores, socioeconomic status and future outcomes* (No. w6943). National Bureau of Economic Research.
- 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).
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control?. *Clinical Psychology Review*, 18(1), 23-45.
- Domitrovich, C. E., Bradshaw, C. P., Poduska, J. M., Hoagwood, K., Buckley, J. A., Olin, S., . . . & Jalongo, N. S. (2008). Maximizing the implementation quality of evidence-based preventive interventions in schools: A conceptual framework. *Advances in School Mental Health Promotion*, 1(3), 6-28.
- Dunst, C. J., Bruder, M. B., Trivette, C. M., & Hamby, D. W. (2006). Everyday activity settings, natural learning environments, and early intervention practices. *Journal of Policy and Practice in Intellectual Disabilities*, 3(1), 3-10.
- Dunst, C. J., Trivette, C. M., McInerney, M., Holland-Coviello, R., Masiello, T., Helsel, F.K., & Robyak, A. (2008). Measuring training and practice fidelity in capacity-building scaling-up initiatives. *CELLpapers*, 3(1). Retrieved from: http://www.earlyliteracylearning.org/cellpapers/cellpapers_v3_n1.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.
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, 18(2), 237-256.
- Fixsen, D. L., & Blase, K. A. (2008). *Drivers framework*. Chapel Hill, NC: The National Implementation Research Network, Frank Porter Graham Child Development Institute, University of North Carolina.

CHAPTER 8 AN OVERVIEW OF IMPLEMENTATION RESEARCH AND FRAMEWORKS IN EARLY CARE AND EDUCATION RESEARCH

- Fixsen, D., Blase, K., Metz, A., & Van Dyke, M. (2013). Statewide implementation of evidence-based programs. *Exceptional Children, 79*(2), 213-230.
- Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core Implementation Components. *Research on Social Work Practice, 19*(5), 531-540.
- Fixsen, D. L., Blase, K. A., Timbers, G. D., & Wolf, M. M. (2001). In search of program implementation: 792 replications of the Teaching-Family Model. In G. A. Bernfeld, D. P. Farrington, & A. W. Leschied (Eds.), *Offender rehabilitation in practice: Implementing and evaluating effective programs* (pp. 149-166). New York: Wiley.
- Fixsen, D., Naoom, S., Blase, K., Friedman, R., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa: Louis de la Parte Florida Mental Health Institute, National Implementation Research Network, University of South Florida.
- Gormley, W. T., Jr., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology, 41*(6), 872-884.
- Gormley, W. T., Jr., Phillips, D., & Gayer, T. (2008). Preschool programs can boost school readiness. *Science, 320*(5884), 1723-1724.
- Halle, T. G., Forry, N., Hair, E., Perper, K., Wandner, L., & Vick, J. (2009). *Disparities in early learning and development: Lessons from the Early Childhood Longitudinal Study-Birth Cohort (ECLS-B)* [Research brief]. Child Trends.
- Halle, T. G., Paulsell, D., Daily, S., Douglass, A., Moodie, S., & Metz, A. (2015). *Implementing parenting interventions in early care and education settings: A guidebook for implementation* (OPRE Report 2015-94). Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Hartmann, A., & Linn, J. F. (2008). *Scaling Up: A framework and lessons For development effectiveness from literature and practice* (Working paper 5). Wolfensohn Center for Development at Brookings.
- 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.
- Metz, A., Bartley, L., Ball, H., Wilson, D., Naoom, S., & Redmond, P. (2015). Active implementation frameworks (AIF) for successful service delivery: Catawba County child wellbeing project. *Research on Social Work Practice, 25*(4), 415-422.
- 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.
- Meyers, D. C., Durlak, J. A., & Wandersman, A. (2012). The quality implementation framework: A synthesis of critical steps in the implementation process. *American Journal of Community Psychology, 50*(3-4), 462-480.
- Munter, C., Wilhelm, A. G., Cobb, P., & Cordray, D. S. (2014). Assessing fidelity of implementation of an unprescribed, diagnostic mathematics intervention. *Journal of Research on Educational Effectiveness, 7*(1), 83-113.
- Peters, D. H., Adam, T., Alonge, O., Agyepong, I. A., & Tran, N. (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>
- Phillips, D., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., & Weiland, C. (2017). Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects. A consensus statement. In K. Dodge (Ed.), *Issues in Pre-Kindergarten Programs and Policy* (pp. 19-30). Washington, DC: The Brookings Institution.

- Pianta, R. C. (2013). Classroom management and relationships between children and teachers: Implications for research and practice. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of Classroom Management: Research, Practice, and Contemporary Issues* (pp. 695-720). Abingdon, UK: Routledge.
- Reardon, S. F., & Portilla, X. A. (2016). Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry. *AERA Open*, 2(3), 1-18.
- 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>
- Supplee, L. H., & Metz, A. (2015). Opportunities and challenges in evidence-based social policy. *Social Policy Report*, 28(4), 1-16.
- Vavrus, F., & Bartlett, L. (2006). Comparatively knowing: Making a case for the vertical case study. *Current Issues in Comparative Education*, 8(2), 95-103.
- von Hippel, P. T., Workman, J., & Downey, D. B. (2018). Inequality in reading and math skills forms mainly before kindergarten: A replication, and partial correction, of "Are schools the great equalizer?" *Sociology of Education*, 91(4), 323-357.
- Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, 84(6), 2112-2130.
- Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 71(1), 136-149.
- Yoshikawa, H., Leyva, D., Snow, C. E., Treviño, E., Barata, M., Weiland, C., . . . & Arbour, M. C. (2015). Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes. *Developmental Psychology*, 51(3), 309-322.

SECTION 3, CHAPTER 9

**DESIGNING IMPLEMENTATION
RESEARCH TO GUIDE THE SCALE-UP
OF EFFECTIVE EARLY CARE AND
EDUCATION ACROSS SETTINGS**

Michelle Maier, Ph.D., MDRC

JoAnn Hsueh, Ph.D., MDRC



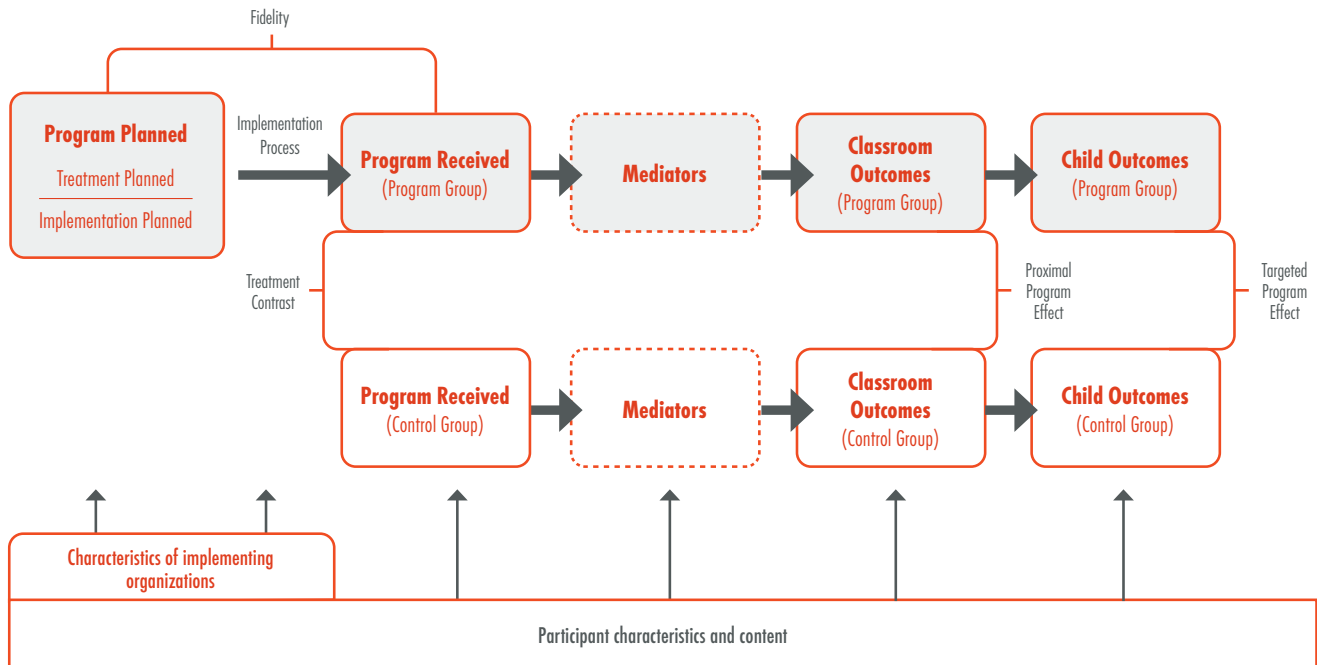
Well-designed implementation research is the key link between small-scale early care and early childhood education (ECE) programs that have been proven to work and large-scale adaptations across populations and settings. Waiting years to see whether programs work provides too little information too late. Ongoing, well-designed implementation research, however, can provide real-time feedback on necessary program adjustments, identify the supports needed to successfully put these programs into action in varied localities and contexts (Martinez-Beck, 2016), and address why and how a program works and under what circumstances. Such research gives the field the information it needs to bring promising programs to wider populations, enabling all children to have access to high-quality learning experiences (Phillips et al., 2017).

This chapter aims to help design strong implementation research to complement rigorous evaluation of ECE programming. It, therefore, has two goals: to provide a set of frameworks to help guide the empirical study of program implementation in an evidence-building context and to discuss potential methodological and measurement problems to consider when taking such an approach. It does not tell developers, researchers, and practitioners what potential areas of inquiry to prioritize in their implementation research. Instead, we aim to illuminate underexplored opportunities and methodological approaches that readers can consider and then apply in their own work. We draw on examples of innovative methodological and measurement strategies from three studies that integrate implementation research into their evidence-building efforts. In doing so, we aim to highlight research opportunities that, by going beyond describing program impacts, can further knowledge and offer a systematic guide to how policy can support at-scale ECE programs that reduce inequities in learning opportunities and disparities in children's outcomes.

A CONCEPTUAL FRAMEWORK

To empirically study program implementation in an evidence-building context, we begin with a conceptual framework for research that examines variation in program effects. Figure 1 outlines the pathway from program implementation to outcomes for ECE centers randomly assigned to receive a program (program group) and those assigned to proceed with business as usual (control group) (Weiss et al., 2014). Using an example program of a new curriculum combined with teacher professional development, researchers often hypothesize the following theory of change: the new program leads to improvements in classroom outcomes (such as more and better instruction) and ultimately to improvements in children's outcomes. Researchers may also propose a set of hypothesized mediators, such as increased teacher knowledge, more positive attitudes and beliefs, or improved teacher practices. Figure 1 illustrates this causal pathway of change as well as other critical aspects of implementation.

Figure 1. Conceptual framework for research examining variation in program effects.



Note: Adapted from Weiss, Bloom, & Brock (2014).

The far left of the framework shows that the program that is planned by developers influences the program received by classrooms with and without access to the program (program group vs. control group). The planned program includes the core components and practices for the new curriculum plus the implementation plan needed to put the program in place (e.g., staff professional development such as training and coaching, technical assistance, and other administrative supports). The procedures, methods, or activities necessary to foster implementation of core components and enact the implementation plan is referred to as the “implementation process.” The relationship between the planned program and what is received by teachers and children is described as “fidelity of implementation.” The line between the program received by the program group and the program received by the control group is termed the “treatment contrast,” which is the difference between the average treatment received with and without access to the program.

Along the bottom of Figure 1 are two boxes representing factors that influence or moderate the specified causal relationships. The top box represents staff and organizational characteristics, which are typically hypothesized to moderate many aspects of program implementation. The bottom box represents characteristics of children within the implementing organization and the organization’s social, physical, economic, financial, and political context. These characteristics are typically thought to moderate the whole chain of events from the implementation process to its effects on outcomes and, in particular, the extent to which income, immigrant, racial, ethnic, linguistic, and cultural backgrounds might affect outcomes.

This framework highlights where sources of variation may be likely to influence program effects and, therefore, underscores where research can focus. This includes operationalizing and measuring:

- fidelity of implementation of the program and implementation plan;
- proximal sources of variation in program effects such as treatment contrast, participant characteristics, and program context;
- distal sources of variation such as characteristics of the implementing organization and of the larger system; and
- potential moderators of these relationships.

In the next section, we further describe what may constitute these sources of variation and how they may be studied.

PROGRAM DEVELOPMENT, IMPLEMENTATION, AND EFFECTIVENESS IN AN EVIDENCE-BUILDING CYCLE

Evidence of program and policy effectiveness arises within an iterative cycle of program implementation, adaptation, and evidence-building activities. The process is often conceptualized as beginning with a program model in an early stage of development (pre-scale-up) that is piloted on a small scale and/or in a relatively controlled setting (for example, under the direct supervision of its developers and with eager volunteer participants). The goal at this stage is to clarify and, if necessary, refine the program goals, target population, and key activities and components as they are being implemented. At this stage, accompanying evidence-building activities designed to evaluate programs commonly entail feasibility studies, demonstrations, pilot assessments, and early efficacy tests.

If early efficacy trials establish evidence of effectiveness when the program is delivered on a relatively small scale, the program may move from the promising to the effective stage. At this stage, efforts typically focus on replicating prior results and/or expanding the program so that it can be tested in more diverse populations and contexts. This undertaking, termed “horizontal scaling,” aims to extend services to a small number of sites (Dunst et al., 2006; Hartmann & Linn, 2008). Accompanying evidence-building commonly entails random assignment efficacy trials through which the program is compared to a business-as-usual comparison/control group. Researchers may therefore adjust the goals for program development, moving on to probing under what conditions, across what contexts, and with what populations the program can be expanded, while also seeking to further test the program’s effectiveness.

As the program continues to mature, it is often scaled more extensively, with the explicit goal of building the level of effectiveness evidence for incorporating the program into an existing system to ensure longer-term sustainability, termed “vertical scaling” (Dunst et al., 2006; Hartmann & Linn, 2008). Evidence-building at this point can thus turn

toward testing and mapping systems, infrastructure, and the institutional supports needed to sustain the model across contexts, populations, and conditions.

Embedded in each of these stages of program development are three aspects of evidence-building research (Knox, Hill, & Berlin, 2018; Metz et al., 2016):

- **implementation** of the program model, which is continually in flux and evolving at each stage of program development;
- **adaptation** of and adjustment and improvement to the defined program model, organizational and system supports, and infrastructure; and
- **building** impact evidence by testing the program model.

In essence, these evidence-building activities have a cyclical relationship; iterative feedback loops aim to strengthen the model as the circumstances, context, and environment in which the program is being delivered evolve, which in turn can help the program operate successfully at each new stage of program development (Knox et al., 2018).

ECE can benefit by aligning implementation research designs and measurement to this evidence-building cycle and stages of program development. As Manno and Miller Gaubert (2016) argue, (a) many implementation research topics and questions are relevant across stages, but depending on whether a program is undertaking horizontal or vertical scale-up, the specific research questions and their emphasis will be slightly different; and (b) even in early stages of program development, implementation research can lay important groundwork for informing future scale-up.

For instance, applicable evidence-building activities in later stages of program development include large-scale studies of evidence-based programs or practices that have expanded widely. Such studies allow researchers and policymakers to examine the effectiveness of these programs across a broader set of contexts, populations, and locations. This type of study has become more prevalent; examples include the Early Head Start Research and Evaluation Study (Administration for Children and Families, 2002), the Head Start Impact Study (Puma et al., 2010), and the Mother and Infant Home Visiting Program Evaluation (Michalopoulos et al., 2015). They provide unique opportunities for researchers to rigorously ascertain sources of variation in program impacts by taking advantage of the multisite designs of such studies, which would not have been feasible in earlier stages (e.g., Weiss et al., 2017). For example, in a secondary analysis of the Head Start Impact Study (Puma et al., 2010), Bloom and Weiland (2015) found substantial variation in impacts generated across sites—variation that suggested Head Start may be more effective when fewer ECE options are available across locations and for dual language learners and Spanish-speaking children. But these kinds of multisite and national evaluations are relatively rare, even though they create unique opportunities to explore variation in the way organizations adapt components of the model and in

intervention fidelity across providers, contexts, or populations. As a result, we have relatively little information about how program models can maintain or even increase their effects as they are widely implemented.

Additionally, even at earlier stages, researchers are presented with opportunities to examine drivers of implementation that can directly or indirectly influence program effectiveness, and the results of such examinations can be useful in addressing scale-up questions of interest (Fixsen et al., 2005). Often, in early stages of program development, less systematic data are captured about more indirect drivers of implementation. However, it is nevertheless helpful to situate the program from this perspective, because these factors become influential sources of variation in implementation and impacts as programs are tested further and scaled. Thus, these topics serve as organizing tools that help researchers explore areas of inquiry for implementation research. The helpfulness of the information yielded from such studies also makes the case for more systematic data collection on these factors and for broadening the conceptualization of measures and research designs that aim to address questions at different stages of development. In undertaking this research, we may be able to build a more systematic body of evidence that can be used to ensure effective, high-quality ECE at scale that improves learning and developmental outcomes for a diverse population of young children.

ADVANCING ECE IMPLEMENTATION RESEARCH: MEASUREMENT AND METHODOLOGICAL CONSIDERATIONS

► Potential methodological approaches in implementation research

Incorporating a strong implementation study in ECE evaluations is necessary for understanding the why behind the effectiveness (or lack thereof) of a program and how best to bring a program to scale. But implementation studies can take multiple forms, using quantitative, qualitative, or mixed-method approaches. Quantitative efforts are more objective, closed-ended, and numerical in nature; use statistical analysis; and commonly rely on methods like surveys, direct assessments, structured observations, and administrative data. Qualitative efforts are more exploratory, subjective, and open ended in nature and typically rely on one-on-one interviews or focus groups (conducted at a single time point or multiple time points), ethnographies, document reviews, unstructured or semi-structured observation, and case studies, among others. Quantitative approaches in implementation research try to quantify constructs of interest—such as the level of fidelity achieved; participants’ attitudes, competencies, and behaviors; and the degree of service contrast observed. In contrast, qualitative approaches may try to explore what underlies participants’ attitudes, competencies, and behaviors as well as their perspectives on how and why fidelity or a service contrast was achieved. Mixed-method approaches combine these two types of methods.

Each approach has notable strengths and weaknesses. The quantitative approach allows us not only to assess the direction and magnitude of relationships among constructs of interest but also to compare the magnitude of

such relationships for different subgroups and to compare the results with those of prior studies that used the same measures with similar populations. Quantitative data also can be captured with larger samples at lower costs than qualitative data, therefore, making such data potentially more generalizable to the population of interest. But the downside to quantitative data is that the constructs of interest need to be prespecified, operationalized, and measured and that measures of these constructs must have been validated for or deemed reliable with the population of interest.

The qualitative approach has the potential to capture rich, descriptive information about people's behaviors, attitudes, perceptions, and experiences as they unfold in contexts that are changing as a function of new policy and programmatic efforts. Further, the often exploratory, inductive, and open-ended coding process of most qualitative studies allows researchers to begin to delineate a series of transactional and dynamic processes in settings that are often difficult to capture with more standard quantitative measurement approaches and thereby develop a theory. At the same time, qualitative approaches do have limitations. Most qualitative implementation studies are fairly small in scale due to the costs of collecting and analyzing qualitative data. Most rely on samples of convenience, developed through snowball strategies. Findings and emergent theories developed with narrow samples require replication and further investigation if researchers are to understand the extent to which the processes identified might be relevant to broader populations and other contexts.

Balancing the strengths and limitations of different methodological approaches in the context of large-scale ECE implementation research can be challenging. We often see focused qualitative endeavors added on to larger-scale implementation and evaluation studies that rely primarily on quantitative data sources. Focusing on a narrow question with qualitative data collection within the scope of a broader implementation or evaluation study provides a unique perspective through which to assess the experiences and perceptions of staff or participants involved with the program or policy initiative and can shed light on and contextualize the findings of the broader study.

► **Topics of inquiry in implementation research**

Drawing on the conceptual framework put forth by Weiss et al. (2014), in this section we highlight six main topics of inquiry for the study of program implementation:

1. Treatment planned, offered, and received
2. Implementation plan and system supports
3. Characteristics of participants
4. Characteristics of the organization/provider implementing the program
5. Institutional and contextual factors external to the organization/provider implementing the program

6. Strength of the service contrast resulting from the program (i.e., the services available to program participants versus those available to control group members)

For each topic of inquiry, we provide a definition and example research questions. We also identify opportunities and underexplored areas, as well as methodological and measurement considerations, to help advance the field. Throughout the section, we draw heavily on three empirical examples that in different ways illustrate how implementation research is critical for the evidence-building cycle:

- ***Making Pre-K Count (MPC) project***, a randomized controlled trial of an evidence-based preschool math program—Building Blocks (Clements & Sarama, 2007)—for which lead and assistant teachers receive two years of training and coaching. Sixty-nine preschools in public schools and community-based organizations with over 170 classrooms and over 2,500 children throughout low-income neighborhoods in New York City form the basis of the longitudinal study, which builds on a relatively extensive body of efficacy evidence conducted by the program developers (e.g., Clements & Sarama, 2007, 2008; Clements et al., 2011; Hofer et al., 2013) and has sought to build an infrastructure that would make its services a longer-term component of the New York City pre-K and educational system. The study features an in-depth implementation research design and measurement approach using both quantitative and qualitative measures. It aims: (a) to shed light on the results of the study’s impact analysis by describing the fidelity of implementation of the curriculum and professional development models, (b) to explore how the math program was experienced by teachers and children, and (c) to guide potential scale-up and replication of Building Blocks across the city.
- ***Researcher-Practitioner Partnerships (RPPs)*** between researchers and the Boston Public Schools’ (BPS) Department of Early Learning that undergird BPS’s data-driven decision making and help build and strengthen its programming. In a long-standing series of collaborations, the RPPs have produced seminal studies about the effects of the BPS prekindergarten program (Weiland & Yoshikawa, 2013), informed the expansion of the BPS prekindergarten model via a delivery system involving community-based prekindergarten and Head Start centers under the purview of the BPS Department of Early Learning (Yudron, Weiland, & Sachs, 2016), and informed more recent efforts via the Institute for Educational Sciences Early Learning Network—a collaboration among BPS, MDRC, the University of Michigan, and the Harvard Graduate School of Education—to extend curricular and professional development reform outward from prekindergarten to second grade.
- ***New York City Early Childhood Research Network***, a hybrid, collaborative early care and education research consortium of eight mixed-methods implementation studies that cut across public school and community-based prekindergarten programs. The studies are part of New York City’s Pre-K for All

(PKA) initiative, an expansion of full-day prekindergarten across the city’s five boroughs. Each one is led by a different research team and is guided by study-specific aims and questions while being tied together by a shared research agenda and a coordinated, place-based sampling approach. Collectively, the studies aim to unpack the complexity of the PKA initiative’s implementation and scale-up efforts. These studies are grounded in the perspectives of the ECE workforce and illuminate overlooked aspects of implementation, such as how administrators, teachers and other support staff, such as coaches, make use of essential elements of the implementation supports prescribed by the PKA initiative, as well as how the system has allocated supports and resources to better address variation in teachers’ and children’s experiences in the classroom. The consortium is a collaborative among academic researchers with the New York City Department of Education, the Mayor’s Office of Economic Opportunity, the Department of Health and Mental Hygiene, and the Administration for Children’s Services; it has funding from the Foundation for Child Development.

► Treatment as planned, offered, and received

The focus of inquiry in this area is *intervention fidelity*, or the degree to which critical components of the program are delivered as expected, in line with the intended program model. Investigation begins with defining the program model, as well as assessing differences between the intended program model and the program model as delivered and received by participants. Fidelity has a number of dimensions (Dane & Schneider, 1998; Durlak & Dupre, 2008), including:

- **dosage**: an index of the quantity of delivery, also referred to as “exposure” (e.g., how many sessions were implemented? How long did they last? How frequently did they occur?)
- **adherence**: the extent to which the specified program content was delivered as described in program materials and manuals
- **quality of delivery**: a measure of qualitative aspects of the manner in which the program components are delivered
- **program differentiation**: the extent to which a program’s theory and practices are distinguishable from other programs, which is gauged to ensure that participants receive only the planned intervention to which they are assigned
- **participant responsiveness**: a measure of participants’ response to the program (e.g., engagement levels, enthusiasm)
- **program reach**: rate of involvement and representativeness of program participants within the intended/eligible population
- **adaptation**: changes or modifications made to the original program during implementation

For programs in an early development stage, this topic of inquiry often focuses primarily on developing and refining the program model and theory of change. In contrast, later stages of program development tend to focus more on the degree to which intervention fidelity along various fidelity dimensions is achieved. Common research questions include “What program was planned and offered?,” “What program components did children receive?,” and “To what degree was there fidelity to the planned program model?”

For example, the MPC study of the Building Blocks program—a 30-week pre-K math curriculum that targets numeric, geometric, and spatial topics and skills—uses online coach logs to capture how often components (whole group and hands-on math centers that are set up daily and small group and computer activities that children participate in weekly) are delivered, the quality of teachers’ delivery of the components, and the overall quality of implementation for lead teachers. Input from the curriculum developers is used to devise benchmarks to monitor the level of intervention fidelity achieved (Mattera et al., 2017). Collection of such information across the school year allows the researchers to describe intervention fidelity in terms of dosage as the extent to which teachers are able to implement most of the main curricular components successfully at levels prespecified by the research team (Morris et al., 2016). It also highlights which curricular components may be more challenging to deliver (computer activities, in this particular case) and how implementation of those components may have changed over time. Further, qualitative findings show that, overall, teachers report engaging in formative assessment activities and differentiation practices that are highly aligned with the training they received (Leacock et al., 2016).

Answering these types of questions in the early stages of program development can help researchers produce meaningful metrics for assessing fidelity to the original model in future scale-up efforts and can help identify which elements of the program model are most essential, reveal which adaptations are appropriate and effective, and make clear what are reasonable expectations for fidelity—all of which are areas of concern once expansion efforts are underway due to cost and operational considerations. In later stages of program development, opportunities arise to describe the degree of variation or consistency in implementation of the program model across populations, locations, and contexts, as well as to link variation in implementation to variation in program impacts. Furthermore, as we underscore later in our discussion, collecting information on intervention fidelity also becomes critically important across all stages of development, as it helps show how fidelity changes as the program is replicated or scaled and makes it possible to examine the strength of the treatment contrast (Cordray & Pion, 2006; Hulleman & Cordray, 2009), even if adaptations to the original program model are made.

Methodological and Measurement Considerations. Most implementation research in this line of inquiry takes a single point-in-time approach to measurement. For example, commonly used methods for measuring intervention fidelity include checklists, surveys, observations, and interviews that typically capture a hypothesized steady state of operation (often thought to be in late winter or early spring in the context of a school year) (e.g., Preschool Curriculum Evaluation Research Consortium, 2008). Such measurement approaches inherently characterize implementation as a static set of processes.

Repeated measurement strategies and designs, in contrast, allow for exploration of dynamic processes and changes in intervention fidelity over time. Measurement approaches, like time use, daily diaries, or surveys collected on an ongoing basis can illuminate consistency in dimensions of fidelity such as dosage, adherence, and quality, allowing researchers to (a) map the arc of changes in implementation as teachers progress toward achieving fidelity to the intended model, (b) predict the variation in implementation that can be expected at different points in time, (c) show how this pattern might differ across multiple years of implementation as the program model matures, and (d) glean insights into the challenges faced by or adaptations made to the program model (see Odom et al., 2010, and Zvoch, 2009 for examples). Findings from MPC, for example, underscore that it's important to understand the arc of implementation within a given school year and across multiple years. Here, with repeated measures of dosage and quality of curriculum implementation collected across two years, the findings suggest that that dosage of all MPC components dips slightly during the winter holiday season (November–December) and toward the end of the year (May–June), a typically more chaotic time (e.g., field trips, moving-up ceremonies). Yet it appears that two years of professional development help teachers start a second school year strong, both in terms of the amount and quality of curriculum implementation, which has implications for the dosage of the curriculum that children receive over a single year. Notably, the quality trends suggest that the overall level of quality achieved each year does not appear to be very different. This kind of information not only can help set expectations when scaling up Building Blocks and when thinking about how curriculum implementation may change across multiple years of implementation but also can suggest potential hypotheses that can be tested in later research.

Processes that feed into the adaptation and evolution of a program model are also important to measure and describe, as they could be relevant to strengthening program effects (e.g., Cannata & Rutledge, 2017; Center on the Developing Child at Harvard University, 2016; Chambers, Russell, & Stange, 2013). For example, the experiences of those implementing the model arguably can best be captured by the qualitative or ethnographic work of staff that links their experiences of transitions and changes brought about by the program model with changes in their delivery of the model. This could help answer interesting implementation questions such as “What are the staff’s perceptions of the model as it is being rolled out?,” “What personal narratives do teachers supply about the purpose of the model and how its components affect their interactions with children?,” and “What difficulties and successes have teachers had in implementing these components, and how do they intersect with their daily experiences working with other staff and with children?” Research that is taking up these issues includes studies being conducted as part of the New York City Early Childhood Research Network that mix qualitative and quantitative methods to better understand the relationships among characteristics of ECE professionals, program components and supports, and classroom instruction in the midst of scaling up universal pre-K.

Another area of potential study in implementation research is analyzing the transactional processes involved in implementing a new model with fidelity, the results of which can then be used for continuous quality improvement efforts. The evolution of BPS’s prekindergarten programming offers a striking example of how such research is

important. In 2013, BPS began rolling out Focus, a system-wide language, literacy, and STEM curriculum that aligns content and instruction from kindergarten through second grade, with the aim of ensuring that kindergarten teachers build effectively on what children are taught in prekindergarten, that first-grade teachers build on what children learn in kindergarten, and so on. Drawing on extant literature and research, the district hypothesized four key ways in which instruction in kindergarten and beyond could be aligned to build off of an already well-developed prekindergarten model: through the content of instruction, the format of instruction, opportunities to tailor instruction to children's skill levels, and professional development support.

The BPS reform effort used a stepwise rollout across the district, an implementation model where the new curriculum for a given grade level is first piloted and then scaled across the district. Yet while the aligned curriculum was being developed and brought to scale across the district, it was unclear whether teachers were implementing Focus as designed or intended, whether BPS should allocate resources and professional development to support teachers in their implementation of Focus and if so, how, and how to ensure that BPS's decision-making around adaptations to the Focus model supported children's gains in the ways intended. In 2016, a collaborative effort was launched to build a data infrastructure that addresses BPS's desire to support children's growth from prekindergarten through third grade by continuously assessing and improving the curricular model. At the core of this work is the development of fidelity tools, co-constructed by researchers and BPS staff. After various iterations and pilot testing of the program, researchers trained BPS coaches and staff to collect fidelity data using the tools. BPS coaches collected prekindergarten data across 40 schools in 2017, kindergarten data across 53 schools in 2018, and first-grade data across 28 schools in 2019; they are planning to collect second-grade data in 2020.

The fidelity tools are designed to capture not only dosage, adherence, and quality of implementation for a given grade but also a set of intentional teaching practices and classroom interactions that are supported by the curricular model and cut across curricular components. These practices and interactions help to extend children's learning and development of unconstrained, higher-order skills—such as receptive and expressive vocabulary, critical thinking, and problem solving—that are thought to contribute to sustained academic achievement and success over time. The research team and the BPS plan to continue their deep and meaningful engagement and collaboration with the aim of advancing the field through careful examination of practices in one district that is working hard to improve students' prekindergarten to third-grade experiences.

The fidelity tools therefore aim to build BPS's capacity to collect and use data that can help guide decision-making around the aligned Focus model. The goals are to better understand the variation in implementation of the aligned model, beginning with prekindergarten and extending through second grade; to identify elements of the curricular model, including components, format of instruction, and intentional teacher practices that are crucial for supporting children's within-year gains and sustained growth over time; and to identify which elements and constructs of fidelity are clear predictors of children's gains and to share that data with teachers in easy-to-understand reports that can

help them strengthen their practices. The fidelity tools will allow the coaches and staff to develop fidelity reports and accessible data they can use to guide BPS decision making.

Last, a generally overlooked aspect of research in this area is children’s classroom experiences as related to the program model. Most commonly, studies capture intervention fidelity as delivered by the provider and less so variation in children’s exposure within classrooms to the program model. Using a propensity-score approach to predict subgroups of children based on levels of absenteeism, Arbour et al. (2016) found that children in Chile who demonstrated a higher likelihood of being absent benefited less from a pre-K program than those who had a lower likelihood of being absent. These findings suggest that measuring and exploiting this source of variation can help illuminate how dilution of the strength of intervention fidelity might undermine program impacts in future scale-up efforts.

MPC has also examined children’s experiences more deeply via a qualitative study (a field visit and teacher interviews), looking closely at how teachers differentiated instruction. Findings show that teachers vary in their beliefs about children and teaching and that these beliefs appear to be related to the ways they modify lessons for children, particularly those who struggle in math. The most prominent differentiation strategy for children struggling in math, the MPC study shows, is changing the difficulty level of an activity. One teacher describes planning the difficulty level for children in the following way:

We played X-Ray Vision One a few weeks ago, so I always have my notes, and I write down my notes on my sheet, so before I do the game for the week, always on a Sunday, I go and I look and I plan and I see what they did the game before, and I write little notes by their name, like, “Start from six,” because the last time, I saw that they did one to ten. They knew it. They counted on from any number, so I said, “They can move up a little.”

Teachers report giving math tasks that go beyond the skills the children currently demonstrated to children they consider to be excelling in math; however, many teachers express hesitation about challenging children they perceived as struggling. These qualitative findings, which would have been difficult to tease out via quantitative methods, have several implications for the project of scaling up the Building Blocks program and for the field’s understanding of differentiated instruction more generally.

► **Implementation plan and system supports**

The implementation plan outlines how the implementing organizations or providers plan to operate the program. The plan includes procedures and activities necessary for fostering implementation of the program model’s core components and practices, such as changes in staffing, professional development (i.e., training and coaching), and other supports like the purchasing of materials or the building of partnerships with other organizations that will

enable the implementing organization to deliver the program model as intended. Related implementation research questions include describing the prescribed implementation supports that are in place; implementation fidelity, or the extent to which the implementation plan is delivered as intended; plans for reaching targeted participants (such as teachers, directors, coaches, etc.); and plans for outreach to and recruitment of children who are currently participating in ECE programming.

To maximize learning in later stages, implementation research should go beyond describing what the implementation plan is and look at how the plan is enacted and why supports seem to work (or not). Further, when a program is being replicated or scaled, implementation research could outline the variation in implementation plans across different providers operating the program. This could include depicting system-level mechanisms that help ensure fidelity to the implementation plan—for example, what management, staffing, funding, and structure of oversight systems are needed to help maintain the dosage, adherence, and quality of training and coaching across multiple providers or geographic locations.

Methodological and Measurement Considerations. Often when high levels of intervention fidelity are achieved, particularly in small-scale studies, details of the implementation plan and supports—and fidelity to the intended levels of these supports—are glossed over (Powell & Diamond, 2016). Commonly used measures tend to focus on structural features of the implementation supports, such as the amount, dosage, and frequency of training or coaching received by recipients; the components of professional development (e.g., in-person observation, one-on-one or small-group consultation); and mode of delivery (in-person, via technology, or through a combination) (e.g., Hamre et al., 2010; Wasik et al., 2013; Powell et al., 2010).

But it is important to capture a host of other aspects of the implementation plan and supports, including:

- *process or content* features, such as the quality of interpersonal dynamics between coaches and teachers, the mechanisms for modeling and providing feedback to participants, the content of professional development, and teacher responsiveness to supports (see Diamond & Powell, 2011; Landry et al., 2009; Neuman & Wright, 2010);
- the extent to which there are *conflicting messages* in the objectives and information being shared with teachers via the program or elsewhere, which may have unintended, countervailing implications for the successful delivery of the intended program model; and
- *factors that facilitate the quality of professional development supports* provided to teachers, such as the characteristics, credentialing, experience, and/or qualifications that make a coach or trainer effective and the supervisory and support systems, caseload specifications, and trainings that can inhibit or facilitate a coach's or trainer's ability to support the delivery of a program model.

Because no two program models are exactly the same, development of measures and unique observational coding schemes are needed in this area of inquiry. Initial implementation research that takes a qualitative approach to understanding the implementation plan and supports could help guide the development and design of appropriate quantitative measures and coding schemes. Further, this kind of information can help explain variation in implementation and program impacts.

The consortium studies being conducted under the auspices of the New York City Early Childhood Research Network employ a variety of strategies to unpack experiences with formal and informal sources of implementation supports for teachers' instructional practice during the PKA initiative. For example, two studies in the consortium (Bank Street College of Education and the National Center for Children in Poverty at Columbia University) take a focused look at how administrators, as leaders of community-based and public school PKA programs, explain adherence to staff members and how they monitor whether staff members are following regulations and standards. The studies examine issues like teacher engagement in training and coaching, use of assessments and curricula, staff qualifications, and whether administrators' teaching priorities are synchronized with teachers' perceptions and prioritization of instructional activities in the classroom. Another study, by the Institute of Human Development and Social Change at New York University, uses network analysis to describe the nature of teachers' social networks within and across PKA programs through which teachers acquire different types of information and mentoring to support their classroom practices. Yet another study, led by Rutgers University's National Institute of Early Education Research, examines how coaches working in PKA programs use their time and explores their perceptions around their roles as influencers of teachers' ECE practices.

In a separate but related vein, a group of studies by Hunter College aims to take a more focused look at how teaching staff use formative assessment tools tied to specific curricula in their planning of classroom activities and implementation of the curricular models. Another study headed up by the Institute of Human Development and Social Change explores how administrators and teachers use existing data sources, such as CLASS scores collected as part of the PKA initiative, to strengthen instructional quality in classrooms through improved professional development and related efforts.

Taken together, the New York City Early Childhood Research Network studies shed light on the processes by which information about standards and regulations are translated and internalized by teachers. Such information could be particularly informative for the design of initiatives in and outside of New York City that aim to strengthen the scale-up of high-quality practices via the existing roles of administrators, mentors, and other informal implementation support networks.

► Characteristics of participants

In implementation studies, the intended target population and the population that ultimately is recruited, enrolled, and served are both of interest. While research suggests that low-income, racial and ethnic minority, and dual language-learning children benefit more from ECE (Gormley et al., 2005; Magnuson et al., 2006; Weiland & Yoshikawa, 2013), an important question as a program is scaled continues to be whether a program is effective for all children or just subgroups of children. Accordingly, implementation studies in early and later phases of development should focus on how the sociodemographics and other risk factors of the families and children that are recruited, enrolled, and served differ from those of the intended target population for the program.

Methodological and Measurement Considerations. As a program is scaled and expands its reach, it becomes important to consider how the characteristics of the actual participants might change as a result of changes in the number of participants being served, the number of providers/organizations delivering the program, and geography. Understanding how the sample population that is successfully recruited, enrolled, and served differs from the intended target population of the program or the samples of earlier studies can help explain program impacts (or lack thereof), as well as guide adaptations to the program model made in response to these differences. Recent trials of Building Blocks in San Diego and New York City, for example, did not have the positive effect on children’s math learning at the end of preschool that prior efficacy trials of the model had suggested it would (Clements et al., 2016; Morris et al., 2016). A confluence of factors may have contributed to the unexpected results, among them, the fact that the preschools participating in these studies served more Hispanic children than those in earlier efficacy studies.

At the same time, disparities in the quality of the ECE learning opportunities of children of color, dual language learners, and those with immigrant backgrounds may also be relevant very early in children’s educational experiences (see the chapters in this volume by Iheoma Iruka and Linda Espinosa). Multiple factors are likely in play, such as unequal access to high-quality educational opportunities, implicit bias and racial stereotyping, and a lack of culturally responsive practices that may better support children of color in classroom environments. While such factors have long been acknowledged in K-12 educational systems, in ECE settings these issues and processes—and how they may build on each other in synergistic and interactive ways—remain poorly understood because we have very little theory and only a small body of empirical research that addresses these matters. The research that has been carried out so far suggests that certain practices, interaction methods, and activities are in fact either culturally responsive or at least acknowledge the diversity of children’s backgrounds, languages spoken, and cultures in classroom learning activities. This is one potential set of strategies for a strengths-based approach to enhancing the learning opportunities and achievement of young children of color, children who are learning dual languages, and children from immigrant backgrounds. Here, implementation research has the unique capacity to contribute to underexplored areas in policy and program models that may facilitate or contribute to disparities in children’s learning opportunities.

Being able to understand and detail the processes at play when thinking about disparities in children’s learning opportunities requires new measurement techniques and focused inquiry in areas like implicit biases that are less typically assessed in implementation research. Indeed, there is a need to develop measures, protocols, and observational tools that will allow us to better capture dynamic processes as they unfold in classrooms. Such information in turn would help us better understand how ECE curricular models, as well as implementation supports and systems, can abate negative influences like implicit biases in children’s ECE experiences.

The New York City Early Childhood Research Network has carried out a set of studies that focus squarely on understanding variation in the delivery and implementation of PKA programming as a way to support learning among children who are dual language learners or who come from immigrant or underrepresented cultural backgrounds. One study, led by Fordham University, examines variation in institutional practices, level of preparation, and the amount and types of support provided to teachers in PKA programs that have concentrations of children with racially and ethnically diverse backgrounds. Another study, run by the Research Foundation of the City University of New York under the auspices of the City College of New York and Teachers College, aims to describe the variability in levels of instructional quality and strategies used to engage underrepresented families across PKA programs.

This consortium of New York City Early Childhood Research Network studies also takes a more focused look at the diversity of the ECE workforce, exploring how this diversity influences the implementation of PKA programming and the supports that are necessary to foster implementation. One study, led by the Research Foundation of the City of University of New York through the Borough of Manhattan Community College, examines male ECE teachers’ perceptions of and experiences with supports during the implementation of PKA programming, including recruitment and retention activities, professional development, and mentoring. Another study, carried out by the Institute of Human Development and Social Change at New York University, uses administrative data to describe how teacher qualifications are distributed across PKA programs and addresses differences across community-based and public school settings. Taken together, these studies illustrate underexplored ways to illuminate how diversity across a large-scale preschool system influences implementation and children’s learning experiences and opportunities in the classroom.

► **Characteristics of organizations/providers implementing the program**

The credentials, academic qualifications, prior work experiences, attitudes, beliefs, knowledge, teaching priorities, readiness, buy-in, motivation to execute the program model, engagement, and stress and burnout of front-line staff carrying out ECE programs as well as supporting staff such as administrators, directors, trainers, and coaches are commonly captured in implementation studies. Other important constructs include information about the organizational climate and culture, the extent to which the leadership is committed, staff turnover rates, the population served, the governance and staffing structure, funding, and the resources and capacity for taking on and maintaining the program and implementation supports.

Examining staffing, management, and organizational characteristics such as these is critical to understanding implementation success and effectiveness or the lack thereof as the program enters different phases of development. Documenting these characteristics in a systematic way early on can impart operational lessons and help predict the types of adaptations to the program model and implementation plan required or the degree of change in preexisting organizational characteristics needed to successfully support the delivery of the program when scaled. As the program moves toward later stages of development and scale-up and the scope of the reach of the program increases, there will likely be more opportunities to exploit naturally occurring variation in organizational characteristics and thereby further assess the importance of these drivers in supporting or inhibiting a program's success and effectiveness.

The importance of moving toward identification of organizational characteristics, management factors, and other processes within organizations that can support or inhibit program success is underscored in a mixed-methods study conducted by Christina Weiland and her colleagues. This study describes the 2.5-year pilot scale-out of the BPS's prekindergarten model into 14 community-based preschool classrooms in high-poverty areas. Weiland and colleagues collected data on instructional quality in each classroom at baseline and at the end of each school year, conducted interviews with key stakeholders at multiple time points, and measured fidelity of implementation in the second and final year of the pilot. The findings indicate that although use of intervention components was high, by the end of the pilot, intervention fidelity of the curricula was generally low, with the community-based classrooms showing lower levels of instructional quality than their BPS-counterpart classrooms (Yudron et al., 2016). Qualitative data pointed to a number of structural factors in the community-based settings that appeared to interfere with implementing the BPS prekindergarten model with fidelity, such as the flexibility permitted in attendance, the lack of common planning time among teachers, the use of mixed-age classrooms, and higher turnover rates among teachers. All of these highlight the need to attend to structural distinctions among pre-K programming delivery models.

Methodological and Measurement Considerations. As the list of potential factors we have listed suggests, the scope of what could be examined is vast. Yet we know that none of these influences operate in isolation from each other but rather are likely linked with others in predictable ways. Tracking dynamic and interactive changes within settings and across levels of ecological analysis could help advance our understanding of contextual factors and their influence on implementation. Changes at a systems level may require intervening levels of institutional and organizational change to ultimately support implementation of the program model and bring about changes in the classroom as experienced by a child. A new curricular model and professional development supports, for example, could influence and be influenced by not only organizational characteristics but also contextual factors over time. Integrating quantitative and qualitative data can illuminate what changes—across different levels and within the implementing organization—shape how the program is being implemented. Research on these linkages and the patterns of organizational, participant, and—as we describe next—system and contextual influences could help the field identify subsets of factors that are most salient.

Implementation research should go beyond describing what happens in the classroom and also look at the broader set of contextual factors that might influence the nature of classroom interactions among teachers and children.

To the extent that the root of inequities in children’s outcomes lies in disparities in exposure to high-quality, adaptive, and responsive learning opportunities in ECE settings, implementation research should go beyond describing what happens in the classroom and also look at the broader set of contextual factors that might influence the nature of classroom interactions among teachers and children. Indeed, such processes may be embedded in institutional systems and settings—as a result of cultural norms, structural biases in ECE settings, and resource allocation—in a way that promotes inequity in children’s experiences. By investigating whether disparities in classroom experiences are evident, as well as how and why they might persist at an organizational level, implementation

research has a unique opportunity to augment our understanding of the role organizational characteristics may play in furthering inequity and how to address it.

► Contextual factors external to an organization

Investigating the contextual factors external to the implementing organization can help to situate the findings from evidence-building efforts of a program at different stages of development. Contextual factors include the funding and policy environment, rules and regulations, and local economic and population characteristics. In early stages of development, implementation studies can aim to describe the systems or structures that are in place as the program is being delivered. This information can be used to guide decisions about the feasibility of scaling the program to particular locations or to identify key funding and policy changes that would be needed for the program to be successfully scaled. When a program operates on a larger scale, systematically documenting contextual factors can provide an opportunity to learn more rigorously about how variation in contextual factors explains when, where, and how a program is more or less effective.

In the MPC project, for example, the research team is interested in describing the context in which MPC is being implemented: New York City preschool programming. The team has found that the preschool landscape in NYC has changed over the course of the study as various reform initiatives have been rolled out, including the Common Core standards, the EarlyLearn initiative (which links quality early care and education standards to child outcomes and has consolidated funds for child care, Head Start, and pre-K to support quality early care services), and Mayor de Blasio’s Pre-K for All initiative (which expanded the number of full-day pre-K slots). These changing circumstances appear to be a driving force in findings regarding the magnitude of the service contrast in MPC, which ought to be taken into account when scaling the model in other contexts.

Methodological and Measurement Considerations. In other policy domains, analysts have assessed patterns of co-occurrence of select contextual dimensions. For example, in the welfare, anti-poverty, and employment policies adopted in the 1980s and 1990s, several common policy dimensions emerged that varied in their mandatory work requirements and their provision of earnings supplements to help sustain families' incomes, time-limited benefits, and child care subsidies, which brought about differential patterns of increases in family income, child care arrangements, and children's outcomes (Morris et al., 2001; Morris et al., 2005). Taking a holistic approach to capturing a combination of potential influences across ecological levels by aggregating information or using community-level data to characterize constructs at higher levels of ecological analysis, researchers could adopt a similar idea to characterize typologies of ECE systems. They could then sample purposefully with these typologies in mind to analyze how this variation might influence program implementation and what impact it might have on children. For example, Coburn et al. (2016) characterize four policy regimes defined along dimensions of alignment with and accountability to the Common Core Standards with hypothesized differential consequences for instructional practices. Following this model, we may be better able to identify sets of processes with cumulative or countervailing influences that moderate implementation or program impacts or that capture the reciprocal nature of influences across levels of system functioning. Such research could guide when, where, and how to scale effective programs.

A related consideration is how challenging it typically is for researchers to amass a sample in smaller-scale implementation studies that allows them to systematically assess and generalize findings with broader contextual and situational influences in mind. To address this issue, the consortium of studies in the New York City Early Learning Network is using an innovative, coordinated, and place-based sampling approach that cuts across public school and community-based prekindergarten programs. A set of community districts in New York City was stratified by the level of resources available in the community using NYC demographic data and city data. From this, researchers selected nine community districts that reflected NYC demographics and were distributed across low, moderate, and high levels of concentrated households living in poverty. Using publicly available administrative data, they identified an eligible pool of PKA programs that served 4-year-old children across the nine community districts. This pool was then used to identify study-specific samples of PKA programs that were stratified to ensure representation of each community district and setting type (public school and community-based PKA programs), as well as racial, ethnic, and linguistic diversity in student-level characteristics, among others. Thus, the coordinated sampling strategy fulfilled practical considerations by ensuring that the research teams did not overtax participating programs with research activities and that each study had a sufficient sample to fulfill its specific aims. It also furthered the learning agenda by guaranteeing some generalizability across the study-specific samples that could help identify emerging cross-cutting themes and show how community-level characteristics might shape findings across studies.

► Service contrast resulting from the program

The effectiveness of a program is a function of a culmination of two sets of influences: the strength of the critical components of the program model being tested and the degree of service contrast (Hulleman & Cordray, 2009), or the difference in experiences with active ingredients of the program model versus other services that might be available to the target population of the program model. We have thus far delineated influences that strengthen or undermine the quality of a program's services as delivered and received by participants, but strengthening the implementation of a program alone is not sufficient to guarantee positive impacts of these investments in ECE at scale. For example, Mendive et al. (2016) found that teachers in a pre-K program in Chile (Un Buen Comienzo) demonstrated fidelity to teaching practices prescribed by the intervention, which they measured by using videotapes of classrooms at three points during the year to assess dosage and adherence. Yet the levels at which teachers engaged in such practices were only modestly higher in the intervention than in the control group, which may help to explain the overall absence of intervention impacts on children's skills.

The research from MPC underscores the need to examine whether some of the primary services being put in place through the program (e.g., training and coaching; math curriculum; math software; monitoring of student progress in math) were being received in the control group. Understanding the services received by the control group, and the degree to which that differs from the program group, guides analysis of the service contrast. This has proved to be particularly important in the MPC study, which, as noted, coincided with several initiatives aiming to improve the academic quality of pre-K instruction in New York City. The team has found that in control sites, a substantial amount of teacher-led math instruction—about 35 minutes in a 3-hour observation—is being delivered at the end of the second year. That is much higher than reported in control group sites in prior Building Blocks studies (Clements & Sarama, 2008; Clements et al., 2011). Such a high level of math instruction in typical New York City pre-K sites may make it harder to detect the effects of Building Blocks (Morris et al., 2016), highlighting the need to interpret impacts (or lack thereof) while considering the service contrast and larger context of the study.

With that said, the amount of math-related professional development and the use of math curricula do yield a distinct service contrast in MPC between program and control conditions (Morris et al., 2016). Quantitative survey data on math-related services, collected at the end of the second year of implementation from school administrators, showed that teachers in control sites received less coaching in math: 66 percent of control sites reported that their pre-K teachers received no coaching in math, and those that did report some coaching described teachers as receiving far less than the program group did. Lead teachers in control sites were offered about 13.8 total hours of training on math, less than half the 30 total hours of training on math that lead teachers in program sites were offered in the same year. Although many control sites reported using some aspects of a math curriculum, there still appeared to be a service contrast: 42 percent of control sites reported using a published math curriculum compared with 100 percent using Building Blocks in program sites, and about half of the control sites reported having computer software with math activities compared with 100 percent of program classrooms that used Building Blocks math computer software.

Thus, a systematic understanding of the service contrast, over the course of different program stages of development, should be a key goal of any implementation study aimed at optimizing the extent to which programs reliably produce positive impacts for young children. This is particularly important given that prior evidence suggests the magnitude of the service contrast is diluted as programs that begin as hot-house, small-scale studies in controlled settings are replicated and scaled (Hulleman & Cordray, 2009). It is thus critical to reassess the strength of the service contrast as the program is delivered in new contexts and environments and with different populations, especially given the changing landscape of ECE programming. Such information can reveal which aspects of the program model add the most value relative to the mix of services that are already available and help to identify strategies for expanding effective programming to reach a broader number of children across localities and contexts.

Methodological and Measurement Considerations. Researchers can bring service contrasts to light in many ways. For example, they can collect descriptive information about other services in the community. Or they can explicitly measure the services received by teachers or children who are in a control or comparison group and then compare the information to the services received by teachers or children in the program group, as in the MPC study. However, capturing the differential in experiences with critical components and practices of the program model requires innovation in measurement and the creation of intervention and implementation fidelity measures that are not only closely tied to the program model and implementation plan but also broad enough that they can be used to capture activities and practices in the control/comparison condition (for examples, see Hulleman & Cordray, 2009; Preschool Curriculum Evaluation Research Consortium, 2008; Bierman et al., 2008; Mattera et al., 2013). When measuring the service contrast, it is also important to assess not only dosage (the amount of services being received or how often they are received) but also the quality of those services.

CONCLUSION

This chapter aims to guide the design of strong implementation research to complement rigorous evaluation research of ECE programming. It suggests three key considerations developers, researchers, and practitioners should bear in mind when designing an implementation study. First, implementation frameworks can guide implementation study design. Second, these frameworks can help determine which critical areas of inquiry to prioritize so that a better understanding of the full story of a program, regardless of where it lies in terms of program development stages, can be developed. Third, the degree of breadth—and in some areas, depth—of measurement for each area of inquiry prioritized is important. Some topics lend themselves to quantitative approaches via data sources like surveys, observational tools, and direct assessments, while others lend themselves to qualitative approaches that make use of interviews, focus groups, time-use reports, or document reviews. A combination of approaches, or an intentional mixed-method approach, may prove best depending on what is prioritized given the program's development stage.

We do not state how to prioritize the various areas of inquiry. Instead, we conclude with several questions to help developers, researchers, and practitioners reflect on and address these considerations, so that their unique implementation study can be poised not only to strengthen the particular program under investigation but also to generate insights as to how policy can support ECE programs at scale that address inequities in learning opportunities and disparities in children's outcomes:

- At what stage of development is the program under study? What level of evidence has already been gathered?
- Where in the evidence-building cycle is the program under study?
- What areas of inquiry are most critical to examine given the program's current stage of development and evidence base?
- Which areas of inquiry may provide information most useful for developing the program and design and measurement strategies?

In sum, we call for concerted efforts to design and enhance implementation research that aims to better understand variation in implementation and program impacts from multiple and holistic perspectives. Such research could guide the development of policy and practice to support and sustain effective programming that reaches a broad number of children in scale-up efforts.

References

- Administration for Children and Families (2002). *Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start: Vol. 1. Final Technical Report*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Arbour, M. C., Yoshikawa, H., Willett, J., Weiland, C., Snow, C., Mendive, S., ... , & Treviño, E. (2016). Experimental impacts of a preschool intervention in Chile on children's language outcomes: Moderation by student absenteeism. *Journal of Research of Educational Effectiveness*, 9(51), 117-149.
- Bierman, K. L., Domitrovich, C.E., Nix, R. L., Gest, S. D., Welsh J. A., Greenberg, M. T., ... & Gill, S. (2008), Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development*, 79(6), 1802-1817.
- Bloom, H., & Weiland, C. (2015). *Quantifying variation in Head Start effects on young children's cognitive and socio-emotional skills using data from the National Head Start Impact Study*. New York: MDRC.
- Cannata, M., & Rutledge, S. (2017). Introduction to New frontiers in scaling up research [Special issue]. *Peabody Journal of Education*, 92(5), 559-568.
- Center on the Developing Child at Harvard University (2016). *From best practices to breakthrough impacts: A science-based approach to building a more promising future for young children and families*. <http://www.developingchild.harvard.edu>
- Chambers, D. A., Glasgow, R. E., & Stange, K. C. (2013). The dynamic sustainability framework: Addressing the paradox of sustainment amid ongoing change. *Implementation Science*, 8(117), doi: 10.1186/1748-5908-8-117.
- Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks project. *Journal of Research in Mathematics Education*, 38(2), 136-163.
- Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, 45(2), 443-493.
- Clements, D. H., Sarama, J., Layzer, C., Unlu, F., Germeroth, C., & Fesler, L. (2016). *Effects on mathematics and executive function learning of an early mathematics curriculum synthesized with scaffolded play designed to promote self-regulation versus the mathematics curriculum alone*. Manuscript submitted for publication.
- Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal of Research in Mathematics Education*, 42(2), 127-166.
- Coburn, C. E., Hill, H. C., & Spillane, J. P. (2016). Alignment and accountability in policy design and implementation: The Common Core State Standards and implementation research. *Educational Researcher*, 45(4), 243-251.
- Cordray, D. S., & Pion, G. M. (2006). "Treatment strength and integrity: Models and methods." In Bootzin, R. R., & McKnight, P. E. (Eds.), *Strengthening research methodology: Psychological measurement and evaluation* (103-124). Washington, DC: American Psychological Association.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, 18(1), 23-45.
- Diamond, K. E., & Powell, D. R. (2011). An iterative approach to the development of a professional development intervention for Head Start teachers. *Journal of Early Intervention*, 33(1), 75-93.
- 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.

CHAPTER 9 DESIGNING IMPLEMENTATION RESEARCH TO GUIDE THE SCALE-UP OF EFFECTIVE EARLY CARE AND EDUCATION ACROSS SETTINGS

- Dunst, C. J., Trivette, C. M., & McInerney, M. (2006). Scaling up early childhood intervention literacy learning practice. *Center for Early Literacy Learning: Cell Papers*, 1(2), 1-10.
- Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* [FMHI Publication #231]. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute; National Implementation Research Network.
- Gormley, W., Gayer, T., Phillips, D. A., & Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology*, 41(6), 872-884.
- Hamre, B. K., Justice, L. M., Pianta, R. C., Kilday, C., Sweeney, B., Downer, J. T., et al. (2010). Implementation fidelity of MyTeachingPartner literacy and language activities: Associations with preschoolers' language and literacy growth. *Early Childhood Research Quarterly*, 25(3), 329-347.
- Hartmann, A., & Linn, J. (2008). *Scaling up: A framework and lessons for development effectiveness from literature and practice* [Wolfensohn Center Working Paper No. 5]. Washington DC: Brookings Institute.
- Hofer, K., Lipsey, M., Dong, N., & Farran, D. (2013). *Results of the early math project: Scale-up cross-site results*. Unpublished manuscript. Peabody Research Institute, Vanderbilt University.
- Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, 2(1), 88-110.
- Knox, V., Hill, C. J., and Berlin, G. (2018). Can evidence-based policy ameliorate the nation's social problems? *Annals of the American Academy of Political and Social Science*, 678(1), 166-179.
- Landry, S. H., Anthony, J. L., Swank, P.R., & Monseque-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology*, 101(2), 448-465.
- Leacock, N., Kou, A., Strassberger, M., Maier, M. F., & Morris, P. (2016, July). *Set goals to meet the needs of your students: Unpacking differentiated instruction on pre-K classrooms*. Paper presented at Administration for Children and Families National Research Conference on Early Childhood, Washington, DC.
- Magnuson, K., Lahale, C., & Waldfogel, J. (2006). Preschool and school readiness of children immigrants. *Social Science Quarterly*, 87(5), 1241-1262.
- Manno, M., & Miller Gaubert, J. (2016). *A conceptual framework for supporting scale-up in implementation research*. Unpublished manuscript. MDRC, New York, NY.
- Martinez-Beck, I. (2016). 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: Paul H. Brookes.
- Mattera, S., Lloyd, C. M., Fishman, M., & Bangser, M. (2013). *A first look at the Head Start CARES Demonstration: Large-scale implementation of programs to improve children's social-emotional competence: OPRE Report 2013-47*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Mattera, S. K., Morris, P. A., Jacob, R., Maier, M., & Rojas, N. (2017). Designing studies to test causal questions about early math: The development of Making Pre-K Count. *Advances in Child Development and Behavior*, 53, 227-253.
- Mendive, S., Weiland, C., Yoshikawa, H., & Snow, C. (2016). Opening the black box: Intervention fidelity in a randomized preschool teacher professional development program. *Journal of Educational Psychology*, 108(1), 130-145.
- Metz, A., Halle, T., Bartley, L., & Blasberg, A. (2016). The key components of successful implementation. In T. Halle, A. Metz, and I. Martinez-Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 21-42). Baltimore, MD: Paul H. Brookes.

Michalopoulos, C., Lee, H., Duggan, A., Lundquist, E., Tso, A., Crowne, S., ... , & Knox, V. (2015). *The Mother and Infant Home Visiting Program Evaluation: Early findings on the Maternal, Infant, and Early Childhood Home Visiting program: OPRE Report 2015-11*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Morris, P. A., Gennetian, L. A., Duncan, G. J. (2005). Effects of welfare and employment policies on young children: New findings on policy experiments conducted in the early 1990s. *Social Policy Report* 19(2), 1-20.

Morris, P. A., Huston A. C., Duncan, G. J., Cosby, D., & Bos, J. (2001). *How welfare and work policies affect children: A synthesis of research*. New York: MDRC.

Morris, P. A., Mattera, S. K., & Maier, M. F. (2016). *Making Pre-K Count: Improving math instruction in New York City*. New York: MDRC.

Neuman, S. B., & Wright, T. S. (2010). Promoting language and literacy development for early educators: A mixed-methods study of coursework and coaching. *The Elementary School Journal*, 111(1), 63-86.

Odom, S. L., Fleming, K., Diamond, K., Lieber, J., Hanson, M., Butera, G., ... & Children's Success Project. (2010). Examining different forms of implementation and in early childhood curriculum research. *Early Childhood Research Quarterly*, 25(3), 314-328.

Phillips, D. A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., ... , & Weiland, C. (2017). *Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects: A consensus statement*. Washington DC: Brookings Institute.

Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M. L. (2010). Effects of an early literacy professional development intervention on Head Start teachers and children. *Journal of Educational Psychology*, 102(2), 299-312.

Preschool Curriculum Evaluation Research Consortium (2008). *Effects of preschool curriculum programs on school readiness: Report from the Preschool Curriculum Evaluation Research initiative*. Washington DC: National Center for Education Research, Institute of Educational Sciences.

Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., ... , & Spier, E. (2010). *Head Start Impact Study: Final report*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.

Wasik, B. A., Mattera, S. K., Lloyd, C. M., & Boller, K. (2013). *Intervention dosage in early childhood care and education: It's complicated* [Research Brief OPRE 2013-15]. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. http://www.acf.hhs.gov/sites/default/files/opre/dosage_brief_final_001_0.pdf

Weiland, C., & Yoshikawa, H. (2013). The impacts of an urban public prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills: Evidence from Boston. *Child Development*, 84(6), 2112-2130.

Weiss, M. J., Bloom, H. S., & Brock, T. (2014). A conceptual framework for studying the sources of variation in program effects. *Journal of Policy Analysis and Management*, 33(3), 778-808.

Weiss, M. J., Bloom, H. S., Verbitsky-Savitz, N., Gupta, H., Vigil A. E., & Cullinan, D. N. (2017). How much do the effects of education and training programs vary across sites? Evidence from past multisite randomized trials. *Journal of Research on Educational Effectiveness*, 10(4), 843-876.

Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 71(1), 136-149.

Yudron, M., Weiland, C., & Sachs, J. (2016). *BPS K1DS: Piloting the Boston Public Schools' prekindergarten model in community-based organizations*. Boston, MA: Boston Public Schools.

Zvoch, K. (2009). Treatment fidelity in multisite evaluation: A multiple longitudinal examination of provider adherence, status, and change. *American Journal of Evaluation*, 30(1), 44-61.



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, & Koulis, 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

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

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.

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.

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).

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.

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).

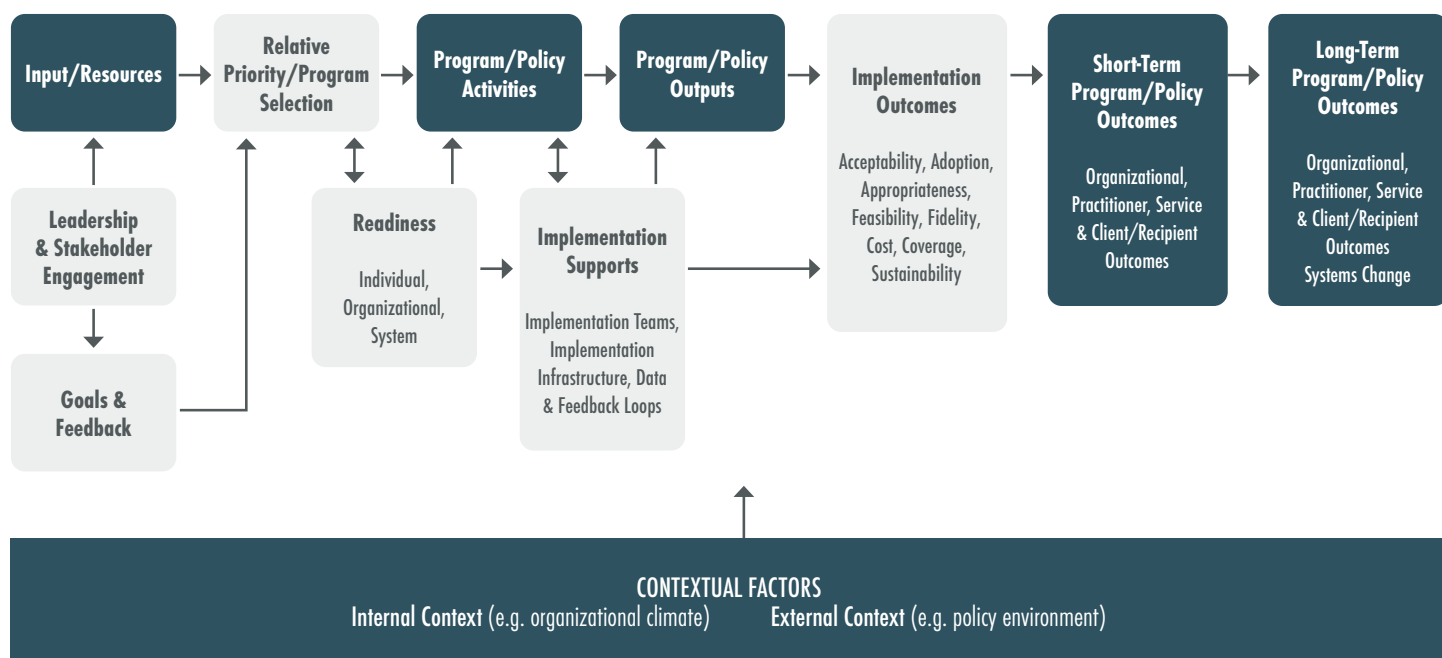
Another contribution that implementation science has made to traditional program evaluation is its treatment of implementation outcomes as distinct from intervention outcomes (Peters, Tran, & Adam, 2013; Proctor et al., 2011). Proctor and colleagues (2011) distinguished implementation outcomes from service 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 CollIN). 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.

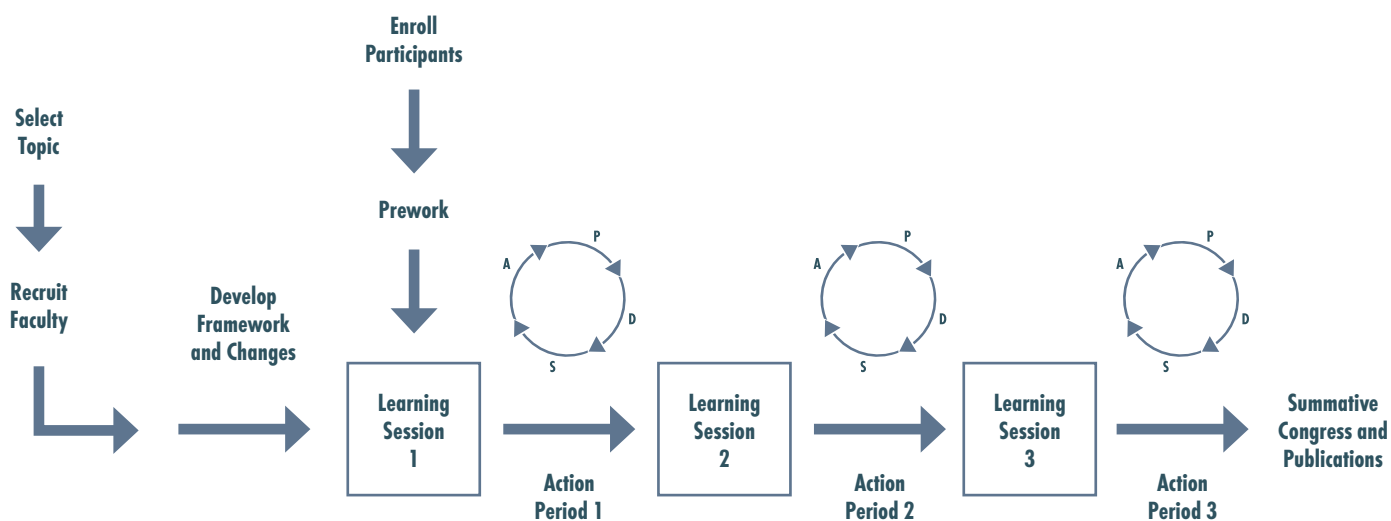
► **Improvement science**

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.

IHI Breakthrough Series Collaborative Model



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 awardees¹⁸ 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 ColIN 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 ColIN 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 ColIN 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 Keese, 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/ptpmrender.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., & Koulisch, 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.

CHAPTER 10 HOW IMPLEMENTATION SCIENCE AND IMPROVEMENT SCIENCE CAN WORK TOGETHER TO IMPROVE EARLY CARE AND EDUCATION

- 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/ccepra_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., Naom, 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.

CHAPTER 10 HOW IMPLEMENTATION SCIENCE AND IMPROVEMENT SCIENCE CAN WORK TOGETHER TO IMPROVE EARLY CARE AND EDUCATION

- 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. *Annals 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., Bunge, 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), 283-288.

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.

CHAPTER 10 HOW IMPLEMENTATION SCIENCE AND IMPROVEMENT SCIENCE CAN WORK TOGETHER TO IMPROVE EARLY CARE AND EDUCATION

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.

SECTION 3, CHAPTER 11

**THE CONTRIBUTIONS OF
QUALITATIVE RESEARCH TO
UNDERSTANDING IMPLEMENTATION
OF EARLY CHILDHOOD POLICIES
AND PROGRAMS**

Sharon Ryan, Ed.D., Rutgers, The State University of New Jersey



INTRODUCTION

The implementation of any educational initiative is a complex endeavor that requires stakeholders to learn new knowledge and skills, apply this learning to their own context, and figure out ways to sustain the reform over time despite changing contextual demands. Much of the implementation research in early childhood education has focused on whether policies or programs work (Weiland, 2018) or whether they are implemented with fidelity. However, implementation is not embodied in a policy or a program—it is the outcome of how groups of people interpret, translate, and practice aspects of policies and programs in particular educational settings (Honig, 2006). As a consequence, innovations vary in how they are implemented, whether they are implemented, and to what extent they are implemented.

Implementation is not embodied in a policy or a program—it is the outcome of how groups of people interpret, translate, and practice aspects of policies and programs in particular educational settings.

In this chapter, I argue that qualitative studies examining implementation of early childhood programs can provide practical information to help policymakers and leaders understand why early childhood programs do or do not fulfill their promise. Qualitative researchers take an interpretive stance, investigating how implementation of an innovation occurs in educational contexts and from the meanings of participants involved in the implementation process (Denzin & Lincoln, 2000). By paying attention to the local and contextual, qualitative research offers a unique position from which to learn about the multiple and conflicting ways innovations go from policy to practice.

CONCEPTUALIZING IMPLEMENTATION RESEARCH

What constitutes implementation research? Theories and perspectives differ, but in this chapter, I use “implementation research” as an umbrella term that encompasses any systematic inquiry of an innovation (e.g., program/intervention/method/pedagogy/policy) in practice, the factors that influence its enactment, and the relations between the innovation, influential factors, and outcomes (Century & Cassata, 2016). Implementation research can examine an innovation vertically (Vavrus & Bartlett, 2006) by how it is taken up and employed at different levels of the educational system (e.g., state, district, and school). Implementation studies may also look horizontally (Vavrus & Bartlett, 2006) at how an innovation is implemented across a number of sites in a range of communities or geographic areas. They can also examine an innovation at different stages of development. For example, in New Jersey a number of quantitative and qualitative studies have been conducted on the state-funded public preschool program, documenting both its impacts over time (e.g., Barnett, Jung, Youn & Frede, 2013) as well as how policy mandates are taken up in local classrooms and communities (e.g., Graue, Ryan, Wilinski, Northey & Nocera, 2018). Because early childhood policies are complex, it is also possible for implementation studies to look

at different aspects of a policy, such as how curriculum models are taken up by teachers (e.g., Ryan, 2004), the approaches of instructional coaches in a program (e.g., Hnasko, 2017; Ryan, Hornbeck & Frede, 2004), or how the higher education system complements a state early childhood policy (e.g., Kipnis, Austin, Sakai, Whitebook, & Ryan, 2013). In this way, implementation researchers can help policymakers adjust aspects of an innovation to achieve improved program quality at the local level.

Implementation has been conceptualized in different ways. The earliest studies in K-12 education tended to look at implementation from a top-down perspective, examining whether a policy or program was implemented as intended or with fidelity (Honig, 2006). This approach tends to view implementation as a technical enterprise in which teachers and other stakeholders accept policies and programs as written and put them into action accordingly. Fidelity studies have often been conducted in early childhood settings when examining the implementation of specific curricula (e.g., Piasta, Justice, McGinty, Mashburn, & Slocum, 2015). Though many policymakers aim to achieve fidelity to implementation when scaling up a particular approach to early childhood programming, not all communities or teachers are willing to implement an initiative as intended, leading to other ways of conceptualizing implementation.

One such way derives from school reform studies of state interventions, such as the Rand Change Agent Study (Mclaughlin, 1987). These studies tended to show that implementation on a large scale was a matter of mutual adaptation as teachers and leaders altered policies and programs to fit their contexts. This perspective assumes that there will always be some adaptation of innovations and researchers should therefore pay attention to whether and how innovations are taken up and what these adaptations look like in practice. Implementation science attempts to do this by developing logic models that identify the various levers and contextual factors that might shape or constrain how an innovation is implemented, as well as the relations between differing aspects of an innovation and how these might lead to expected outcomes. This conceptualization of implementation rests on the assumption that although the policy or program may be changed a little, those doing the implementing will follow the intent of the innovation.

More recently, implementation researchers have begun to theorize about implementation as *enactment*—a complicated network of relations that assumes the movement from innovation to practice is multidirectional, not just top down or bottom up, as well as deeply political (Datnow, 2006; Honig, 2006). From this perspective, the implementation process is influenced and shaped by many agents (from children to policymakers) with varying levels of power and influence within educational settings that constitute a nexus of multiple policies at any one time. Researchers working from an enactment perspective look at the politics of innovation, and how a wide range of stakeholders working in various networks resist, transform, and implement policy depending on organizational ethos and resources, professional theories, and perceived need (Braun, Maguire, & Ball, 2010).

QUALITATIVE STUDIES OF IMPLEMENTATION

Qualitative or interpretive research is interested in how individuals construct their social worlds and how those worlds are mediated by context and culture (Glesne & Peshkin, 1992). Research from this perspective typically involves spending a lot of time in educational settings, observing and talking with participants to develop an understanding and interpretation of educational phenomena. Qualitative researchers interested in implementation therefore examine innovations in sites of practice, often observing what takes place in schools and early childhood settings; they also shadow key stakeholders (leaders, teachers, families, state-level policymakers, coaches, etc.) and question them about an innovation and the reasoning behind their approach to implementing it. Using both the mutual adaptation and the enactment perspective, this research tends to focus mostly on the implementation of various public policies guiding prekindergarten or preschool.

► Qualitative studies of the implementation of public preschool

Qualitative studies of preschool programs are not new. Early ethnographic studies (e.g., Lubeck, 1985; Lubeck Jessup, deVries & Post, 2001; Tobin, Wu & Davidson, 1989; Tobin, Hsueh, & Karasawa, 2009) examined teaching in local sites of practice in the U.S. and elsewhere to illustrate how different values shaped what preschool looked like in action. These small comparative case studies provided some sense of how local actors and community values mediate practice, but they did not look at the findings in relation to bigger policy issues like program improvement across a multitude of sites. However, investments in public preschool have catalyzed a new genre of policy-capturing studies that tend to look at public preschool implementation at the local level of classrooms and school districts.

► Implementation at the local level

Qualitative studies of implementation at the local level are most often conducted in classrooms, examining preschool teachers' experiences and perspectives of a particular policy (such as a curriculum requirement) or, more broadly, what state or district preschool policy looks like in action. Most researchers employ a case study methodology using multiple data sources (interviews, documents, and field notes) to describe life in preschool classrooms. Some studies look not only at classrooms but also at how preschool is embedded in a district and community. In this way, they illustrate the interplay among the various stakeholders who are trying to create public preschool in a particular location. Such studies shed light on the tensions that arise when school districts partner with community providers to enact preschool systems, as well as the factors that mediate implementation.

Tensions between prekindergarten and K-12. The expansion of public preschool has brought changes to the landscape of early childhood services. In most states, oversight of preschool has transitioned to departments of education (Jacoby & Lesaux, 2017), which in the past were not typically responsible for the education of 3- and 4-year-olds. Many states are also using a mixed service-delivery system in

which preschool is offered through a partnership between local education authorities and traditional service providers such as Head Start and child care sites. Though it is logical to work with experienced providers, a number of qualitative case studies of preschool policy implementation have examined what happens when preschool teachers from different auspices begin to work within this new preschool-to-12th grade system.

People who work with children under five years old often operate with different philosophical and instructional goals than those who teach in K-12. For example, they tend to emphasize that knowledge of young children's learning and development—or what is often termed developmentally appropriate practice—should be the starting place for curriculum and instruction (Copple & Bredekamp, 2009). In contrast, K-12 education tends to focus on subject matter, resulting in more didactic and teacher-led instruction. While this dichotomy is problematic in itself, several studies (Brown, 2009; Brown & Gasko, 2012; Desimone, Payne, Fedoravicius, Henrich & Finn-Stevenson, 2004; Graue, Ryan, Norcera, Northey & Wilinski, 2016; Wilinski, 2017) have examined the clash of values that occurs when preschool teachers start to work with their K-12 colleagues.

For example, Brown (2009) conducted a case study of one large urban district where the prekindergarten teachers worked with administrators to develop an assessment system for 4-year-olds to inform kindergarten teachers. The new assessment tool infused developmentally appropriate indicators in six academic areas (such as language arts, math, etc.) aligned with the state's prekindergarten guidelines, and teachers were encouraged to assess children's learning along a four-point scale using anecdotal records. However, observations, plus interviews with key stakeholders after the first year of implementation, illustrated the tension that arose between the prekindergarten teachers' views of teaching and assessment and that of their elementary school colleagues. Elementary stakeholders argued that the tool did not imbue high academic expectations and that it was not clear how these developmentally appropriate indicators would ensure children had the necessary knowledge and skills for success in kindergarten. Though the prekindergarten administrators and teachers had hoped the assessment tool would facilitate more alignment of child-centered practices, the tool was eventually revised to embody more explicit attention to the content knowledge and skills 4-year-olds must acquire before entering kindergarten. Brown concludes that some of the tensions that arose in this case occurred because district resourcing was tied to third-grade test scores. Therefore, leaders believed it was more important to achieve academic alignment across the P-12 system by focusing on content rather than children's development.

Several other case studies have looked at the tensions between preschool and the K-12 system from the perspective of standards. Standards-based reform began in earnest in the K-12 sector with the No

Child Left Behind Act (2002), which held states accountable for student learning, school progress, etc., according to their standards. The expansion of publicly funded preschool in the U.S. also led to a standards movement. Since 2009, all 50 states have had early learning standards about what young children are supposed to know. Several case studies (Brown, 2010; Graue, Wilinski & Nocera, 2016; Graue, Ryan, Nocera, Northey & Wilinsky, 2016) of prekindergarten have asked: Which standards guide the work of teaching and learning?

Graue et al. (2016) conducted a multi-site case study of prekindergarten implementation in two states: Wisconsin, where programs are locally controlled, and New Jersey, where programs are highly regulated by the state. By observing classrooms in each state over the school year, and through interviews with teachers and administrators, the researchers found that although each state had early learning standards, most prekindergarten teachers felt they had no choice but to align at least part of their curriculum and teaching with K-12 standards by incorporating more instruction in academic content. For example, in one district in Wisconsin, a prekindergarten teacher in a public school was told that in the upcoming year she must use a math curriculum that was designed for 5- and 6-year-olds. In New Jersey, a Head Start teacher reported that the administration expected teachers to infuse more literacy into the Creative Curriculum to ensure that children were ready for kindergarten. To do this, she would bring small groups of children together to work explicitly on key skills during center time, and each week in large group time they focused on a new letter of the alphabet. Therefore, regardless of the policy standards context, it seemed that in these classrooms teachers felt pressured to address K-12 content standards by altering some of their more student-centered practices that were reflected in their respective state's early learning standards.

Looking across these studies, it is possible to see the curriculum and instructional challenges as school districts and community-based providers partner to provide preschool in a particular location. Tensions often stem from the neoliberal discourses shaping education as a whole (Brown, 2015, Graue et al., 2016). With the emphasis on accountability as children move through the school system, both preschool teachers and their elementary counterparts feel particularly pressured to ensure that young children will succeed academically, as measured on academic tests. As a consequence, the research in this area suggests that it is preschool teachers who are shifting their practices to be more in alignment with the demands of the K-3 grades.

The findings from this group of studies suggest that policymakers and leaders of preschool implementation efforts need to consider how to bring key stakeholders together in the initial phases of a program to learn about each other's understanding of preschool, and to try to reach some consensus about the purposes of preschool and what it should look like in action.

FACTORS MEDIATING IMPLEMENTATION

Implementation of any educational reform is mediated by a number of organizational factors (Fullan, 2001). A handful of qualitative studies that look beyond the classroom to examine relationships between teachers and the organizations and the communities in which they work provides some insight into the factors that shape preschool implementation in local settings. In general, these factors tend to be related to resourcing and leadership.

Resourcing. In K-12 education, public schools in a district are governed by a similar funding formula. But Head Start and community-based providers have traditionally received less funding than public schools. Moreover, some services, such as Head Start, receive public dollars, while private for-profit or nonprofit child care centers tend to rely on parent fees. To be sure, in mixed service-delivery systems, child care sites and Head Start supplement their funding with state prekindergarten dollars. However, the limited funding of child care and Head Start sites often means fewer opportunities for professional development, mentoring, planning time, etc., for teachers, as well as limitations when it comes to facilities, among other things (Whitebook & Ryan, 2011). Some qualitative studies have observed how these funding differences play out in teachers' practices and the delivery of quality learning experiences for young children.

For example, in their multi-state case study of six preschool settings, Graue et al. (2018) describe how the resources available in a given organizational context impacted what teachers could do. In New Jersey, where preschool teachers receive equal pay across settings, the auspice shaped how specific routines were enacted. This was most striking with the policy requirement that all children have 45 minutes of outdoor playtime. In two of the prekindergarten programs visited regularly, gross motor time was limited because of inadequate facilities—most notably in the Norwood district, where the classroom was part of a Head Start program with no outside play space. A room had been converted to a gross motor area that included an indoor slide and various equipment like stilts and balls. However, there was no consistent schedule for using this space, in part because on some days adult-to-child ratios could not be met because of limited funds for substitute teachers, and assistant teachers were moved around to meet ratio requirements in various rooms. As a consequence, the schedule for physical play was constantly changed. Celia, a prekindergarten teacher at this site, explained that “sometimes they would have it in the morning, another day we would have it in the afternoon. The kids are going out of control, they need consistency.”

In her case study of three programs in one Wisconsin district that received funding from the state to implement public prekindergarten, Wilinski (2017) describes the economic costs of creating mixed service-delivery systems. In the district of Lakeville, both half-time and full-time programs could apply for 4K funding to offer half-day public preschool. Though they received a per-student rate from the state,

districts determined locally how much funding partner sites would receive. Thus at many sites that had anticipated public funds to offset costs, the reimbursement offered by the district was not enough. As a consequence, some sites lacked the funds to purchase materials or find appropriate substitute teachers, given that the state required qualified teachers in prekindergarten. Complicating things further, schools offered transportation for preschool children to their own sites but not to partner sites, limiting access for families who needed wraparound care in addition to a half-day preschool program for their children.

The most compelling difference in resourcing between many early childhood settings and public schools is teacher compensation. Teachers in public schools typically have better benefits and wages than their counterparts working in Head Start and community-based programs (Whitebook, Phillips, & Howes, 2014). In some states, parity is achieved by giving teachers equal pay for similar qualifications regardless of auspice, but in other states, programs receive a particular level of prekindergarten funding, which they may or may not use to equalize wages. Several studies highlight how the differences in

The most compelling difference in resourcing between many early childhood settings and public schools is teacher compensation. Teachers in public schools typically have better benefits and wages than their counterparts working in Head Start and community-based programs.

compensation produce tensions not only between schools and partner sites but also between prekindergarten and kindergarten teachers. For example, Graue et al., (2018) describe how teachers working in partner sites in New Jersey were frustrated because, as a result of belonging to a different union, they were expected to work in more difficult conditions for similar pay but without the same benefits. Similarly, Wilinski (2017) describes how because districts could determine salaries of prekindergarten teachers, there were inequities in teacher compensation depending on where teachers worked. One district, for example, required that prekindergarten teachers in community sites be paid at least 90% of what a public

school teacher with similar credentials earned. Not only was inequitable compensation a problem, but, as Wilinski points out, the child care sites lacked any kind of pathway for teachers to improve their compensation, leading to teacher turnover from child care sites to public schools.

Even when preschool teachers work in public schools, tensions around resourcing can still arise. In a focus group study with 42 teachers (20 preschool and 22 kindergarten teachers) working in four schools involved in a whole-school reform network, Desimone et al. (2004) found that because of union contracts, preschool teachers in a given school were paid salaries closer to those of teachers in child care sites, despite the fact that many had master's degrees. With preschool having been added to public schools with little space, Desimone et al. (2004) also found that kindergarten teachers were wary of

sharing resources like technology with their preschool colleagues. As a consequence, preschool teachers in this study reported feeling a lack of support from their elementary colleagues and uncertainty about their place in the elementary school.

Policy implementation is constrained or enabled by a site's monetary and physical resources. These studies highlight how important it is for policymakers to think about equity for services and teachers when partnering for public preschool. If equity between public school and community-based settings is lacking, then young children may get less access to a high-quality education.

Leadership. Research on school reform initiatives (e.g., Desimone et al., 2004; Fullan, 2001) has illustrated time and again how important school leaders are to any initiative. Principals provide resources and time for teachers to learn about an initiative and to consider how they might implement it in their own classrooms. Effective school leaders also recognize that change takes time, and therefore they help teachers maintain small steps towards implementation. For example, in their interview study with preschool teachers and kindergarten teachers involved in implementing preschool in public schools, Desimone et al. (2004) found that both school principals and district administrators were key to including preschool in their elementary schools. District leaders provided the clout to ensure that principals persisted with the reform, while knowledgeable principals who were committed to the initiative worked hard to get preschool and kindergarten teachers to collaborate.

Few qualitative studies focus solely on early childhood leadership in the implementation of public preschool. Though of late the field has seen a lot more attention given to workforce issues, in general the research on principals, directors of early childhood settings, and other leaders in different parts of the system is limited. Some evidence is available from case studies of preschool implementation in districts (e.g., Brown, 2009; Graue et al., 2018; Wilinski, 2017), which often interview leaders as well as teachers. In general, these studies would suggest that leaders in educational communities shape the resources available to teachers as well as what teachers are expected to teach.

One of the few studies focused solely on leaders was conducted by Whitebook, Ryan, Kipnis, and Sakai (2008), who interviewed 98 Head Start and private child care directors in 16 of the 31 districts offering public preschool in New Jersey about partnering with school districts to provide preschool. Though the directors conveyed that the infusion of money and district resources had been beneficial to their sites, the majority reported struggling with governance issues between policy requirements and those of the auspice in which they worked. For example, different reporting requirements as well as different staff qualifications meant they were constantly trying to keep on top of paperwork and remain positive in an organizational context in which the public preschool teachers were paid more and had access to on-site coaching as well as more professional development opportunities.

Similarly, in a recent mixed methods dissertation study of leaders of state preschool programs, Northey (2018) found that governance was a constant barrier to achieving the goals these leaders had for the program. State leaders said they struggled to have a voice in policy conversations in their state's department of education and therefore had less opportunity to obtain and maintain resources for their programs. Most leaders in this study were early childhood professionals with leadership training, and yet they felt their expertise was undermined as they—like many of the preschool teachers in their state—attempted to bring quality early childhood practices into K-12 education.

Braun et al. (2011) have argued that implementation researchers often fail to recognize that educational settings are sites of multiple policies interacting simultaneously. Whether they look at a director in a school district or a leader at the state level, these leadership studies suggest that public preschool may be a partnership in name but not always in practice. Without some thought by policymakers as to how to bring different levels of the preschool system together, what children experience as a preschool education may vary considerably.

TOWARD A QUALITATIVE IMPLEMENTATION RESEARCH AGENDA

Focusing on the implementation of early childhood programming in local sites of practice and on the perspectives of participants helps us understand whether and to what extent a policy is implemented as intended, makes it possible to see how policies and programs are shaped by context and local actors, and can help with theorizing change and improvements in practice. However, the research base is limited to a handful of studies, and few of these look at implementation across multiple sites, multiple states, or at all levels of the system. The research reviewed in this paper suggests three possible paths toward a more comprehensive, critical, and policy-capturing use of qualitative research to improve the implementation of high-quality early childhood education systems. These include moving beyond classrooms and school districts to investigate multiple levels of the early childhood system, focusing on multiple stakeholders in the early childhood system, and, finally, considering equity.

► Investigating multiple levels of the system

Think about the multiple levels through which early childhood policy takes place within and across states. To date, most qualitative studies focus on the classroom and teachers' implementation of preschool. Some also look at how classrooms are nested within educational sites and, in some cases, how these educational sites interact with local communities. However, the implementation of early childhood programs such as preschool occurs at multiple levels of the system (Paulsell, Austin, & Lokteff, 2013): for example, through infrastructure organizations such as Head Start grantees, through the system of higher education, and through organizations at the state level. In some states, preschool policy entails a number of system-level supports (e.g., coaching and professional development) that also need to be investigated. By qualitatively mapping and documenting the multiple levels and sites in and through

which early childhood policy is implemented, it might be possible to gain some sense of what shapes stakeholders' interpretations of practice and of which aspects of policy get put into practice and why, as well as to map the way policy becomes practice through multiple layers of the system from the top down, the bottom up, and across key agencies and individuals.

To be sure, qualitative mapping in this way would need to focus on the key components of a system, and might need to focus on some critical cases to show differences across the system depending on where a child is and which agencies and stakeholders are interacting around that site. Such work might thus be able to illuminate the politics of enacting early childhood programming in one community versus another and to isolate the factors that contribute to differences in implementation. This kind of work could then lead to more extensive quantitative and mixed methods studies of the implementation of early childhood programs in a state. It might also contribute to the development of tools to help other states and agencies understand the multiple parts of any early childhood system. At the moment, most would agree that the early childhood system is fragmented, and some of the issues around implementation of any policy or program arise from the fact that most stakeholders only know the parts of the system they interact with.

► **A focus on all stakeholders**

A second and related pathway for inquiry is to concentrate a lot more research attention—through interview studies as well as case studies—on the multiple stakeholders who implement early childhood programming. The current qualitative research base on preschool implementation focuses primarily on the preschool teachers who are on the frontlines of implementation. However, the qualitative studies reviewed here all highlight a tension between the values and practices of preschool educators and those working in the K-12 system. We need more extensive investigation of K-3 teachers' beliefs and practices. This focus would help us understand the sources for their approaches to teaching young children and their resistance to what is known about high-quality early education. It would also help us learn what supports they might need to sustain developmentally appropriate yet academically rigorous instruction in the primary grades. If preschool is to achieve its intended outcomes, children need to experience a high-quality education in the early elementary grades. Yet to date there is little research on systems-building work in preschool through third grade, even though some states have initiatives in place.

Implementation of any early childhood program depends on knowledgeable leadership, whether at the state level, in a particular agency, in a school district, or at a local site of practice. Yet there is a dearth of research to help understand what leaders at various levels of the system are doing as they facilitate the implementation of early childhood programming. This line of inquiry is all the more important given that there is no required credential or certification for early childhood leaders; programming specific to early childhood leadership is limited (Goffin & Janke, 2013); and even in the K-12 system, where leaders are expected to have certain credentials, many who are leading P-3 systems building lack knowledge of early childhood education. Future research needs to gather

demographic data on the leaders implementing early childhood programming, their experience and expertise in early childhood education and leadership, and their professional development needs. Another line of inquiry might be to investigate exemplary leaders of program implementation to get a sense of what skills and strategies these leaders use at different parts of the system to support change.

Early childhood programming and systems building is a social construction involving many stakeholders (e.g., coaches, higher education faculty, community members, agency personnel, etc.). To understand where policies and programs either work or go awry, the perspectives and work of other stakeholders are important. Yet because the current qualitative research base suggests that both leadership and the relations between preschool teachers and their primary school counterparts are sources of tension, these seem to be important starting points.

► Issues of equity

Finally, the qualitative research base on implementation indicates that inequities are occurring in current systems of preschool education, and that these may have inadvertent consequences. The first of these inequities is the difference in resourcing and compensation experienced by teachers depending on where they work (Graue et al., 2016), union contracts (e.g., Desimone et al., 2004), or the state policy guiding the programs. Other authors, such as Wilinski (2017), have highlighted how local control of programs in Wisconsin can lead to a lack of access to high-quality preschool programs and resources like busing for families. In other words, despite the rhetoric that participation in a high-quality preschool program can level the playing field for children from disadvantaged backgrounds, it seems that implementation of policies can have unintended consequences that may contribute to children having less than ideal educational experiences.

To date, most research on the implementation of early childhood programming has been on what works and not on what programs look like in action or who benefits and at what cost (Weiland, 2018). Therefore, another line of inquiry is to look at children's experiences in programs and whether those experiences vary by race, class, gender, social class, and languages spoken. Even with targeted programming for students from disadvantaged backgrounds, there is always variation in who gets the most from curriculum and instruction. Qualitative studies with children and families can be particularly informative here, as they can provide detailed accounts of students' lives in early childhood programs by examining the subtle social relationships that take place in classrooms, and whether some children have more opportunities than others for high-quality interactions with teachers and materials.

Along with studies of children's experiences and learning from families about programs, it is also essential to continue exploring inequities across the early childhood workforce and the impacts of differences in compensation, work environments, and benefits (Whitebook, Phillips, & Howes, 2014). If a lack of parity in compensation, benefits, and opportunities for advancement means that educators leave their programs, then the quality of children's

Along with studies of children’s experiences and learning from families about programs, it is also essential to continue exploring inequities across the early childhood workforce and the impacts of differences in compensation, work environments, and benefits.

experiences is lessened. Qualitative interview studies with early childhood educators can help us learn how early childhood policies may lead to retention or turnover and provide insights into effective strategies for building a qualified and stable workforce. With careful sampling, it might be possible to look closely at differences in staffing patterns quantitatively across states, but also to go deeper by eliciting educators’ perspectives on the intersections between policy, their work environments, and their decisions to stay or leave.

CONCLUSION

The early childhood field has assumed for some time that with evidence of best practices, it is possible to scale up and replicate what works in one site to many programs. But implementation research from a qualitative orientation illustrates that what may be evidence-based is often transformed, adapted, or even ignored in local sites of practice. To date, the potential of qualitative studies to guide policy and practice has been limited to a few states and sites, and rarely have the data from these studies been integrated into larger studies of policy implementation in a state. As the field moves away from questions of what works to investigating the implementation of early childhood programs, it will be necessary to bring researchers from differing orientations together to come up with mixed methods designs that look across programs at a macro scale while also employing qualitative studies to go deeply into variations in context and implementation strategies. With more qualitative studies of implementation across multiple sites, it might be possible to identify which local adaptations make sense and which may unnecessarily undermine best practices for young children and those charged with their education.

References

- Barnett, W. S., Jung, K., Youn, M., & Frede, E. (2013). *Abbott Preschool Program Longitudinal Effects Study: Fifth grade follow-up*. New Brunswick, NJ: National Institute of Early Education Research.
- Braun, A., Maguire, M., & Ball, S. (2010). Policy enactments in the UK secondary school: Examining policy, practice, and school positioning. *Journal of Education Policy, 25*(4), 547-560.
- Brown, C. P. (2015). Conforming to reform: Teaching in pre-kindergarten in a neoliberal early education system. *Journal of Early Childhood Research, 13*(3), 236-251, doi: 10.1177/147671X14538602.
- Brown, C. P. (2010). Balancing the readiness equation in early childhood education reform. *Journal of Early Childhood Research, 8*(2), 133-160, doi: 10.1177/1476718X09345504.
- Brown, C. P. (2009). Pivoting a prekindergarten program off the child or the standard? A case study of integrating the practices of early childhood education into elementary school. *Elementary School Journal, 110*(2), 202-227.
- Brown, C. P., & Gasko, J. W. (2012). Why should pre-K be more like elementary school? A case study of pre-K reform. *Journal of Research in Childhood Education, 26*(3), 264-290, doi: 12.1080/02568543.2012.686471.
- 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.
- Copple, C., & Bredekamp, S. (Eds.) (2009). *Developmentally Appropriate Practice in early childhood programs serving children birth through age 8* (3rd edition). Washington DC: National Association for the Education of Young Children.
- Datnow, A. (2006). Connections in the policy chain: The “co-construction” of implementation in comprehensive school reform. In M. I. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 105-124). Albany, NY: State University of New York Press.
- Denzin, N. K., & Lincoln, Y. S. (2000). Introduction: The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The handbook of qualitative research* (2nd edition, pp. 1-32). Thousand Oaks, CA: Sage.
- Desimone, L., Payne B., Fedoravicius, N., Henrich, C. C., & Finn-Stevenson, M. (2004). Comprehensive school reform: An implementation study of preschool programs in elementary schools. *Elementary School Journal, 104*(5), 369-389.
- Fullan, M. (2001). *The new meaning of educational change* (3rd edition). New York: Teachers College Press.
- Glesne, C., & Peshkin, A. (1992). *Becoming qualitative researchers: An introduction*. White Plains, NY: Longman.
- Goffin, S. G., & Janke, M. (2013). *Early childhood education leadership development compendium: A view of the current landscape* (2nd edition). Washington, DC: Goffin Strategy Group. Retrieved from: [http://qrisnetwork.org/sites/all/files/materials/ECE%20Leadership%20Development%20Compendium%20\(May%202013\).pdf](http://qrisnetwork.org/sites/all/files/materials/ECE%20Leadership%20Development%20Compendium%20(May%202013).pdf).
- Graue, M. E., Wilinski, B., & Nocera, A. (2016). Local control in the era of accountability: A case study of Wisconsin preK. *Education Policy Analysis Archives, 24*(60), <https://doi.org/10.14507/epaa.24.2366>.
- Graue, M. E., Ryan, S., Nocera, A., Northey, K., & Wilinski, B. (2016). Pulling preK into a K-12 orbit: The evolution of preK in the age of standards. *Early Years, 37*(1), 108-122.
- Graue, M. E., Ryan, S., Wilinski, B., Northey, K., & Nocera, A. (2018). What guides pre-K programs. *Teachers College Record, 120*(8), 1-36.

CHAPTER 11 THE CONTRIBUTIONS OF QUALITATIVE RESEARCH TO UNDERSTANDING IMPLEMENTATION OF EARLY CHILDHOOD POLICIES AND PROGRAMS

- Hnasko, A. (2017). *Instructional coaching in early childhood education: A comparative case study*. Unpublished doctoral dissertation. Rutgers University.
- Honig, M. I. (2006). Complexity and policy implementation: Challenges and opportunities for the field. In M.I. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 1-23). Albany, NY: State University of New York Press.
- Jacoby, J. W., & Lesaux, N. K. (2017). Policy-based instructional reform in early education: How US Head Start teachers perceive instructional mandates. *International Journal of Child Care and Educational Policy*, 11(9), 2-21. doi: 10.1186/s40723-017-0034-x.
- Kipnis, F., Austin, L., Sakai, L., Whitebook, M., & Ryan, S. (2013). *The state of early childhood higher education in New Hampshire: The New Hampshire Early Childhood Higher Education Inventory*. Berkeley, CA: University of California, Center for the Study of Child Care Employment.
- Lubeck, S. (1985). *Sandbox society: Early education in black and white America*. New York: Falmer Press.
- Lubeck, S., Jessup, P., DeVries, M., & Post, J. (2001). The role of culture in program improvement. *Early Childhood Research Quarterly*, 16(4), 499-523.
- Northey, K. (2018). *Leading from the top: A study of state early childhood systems leaders*. Unpublished doctoral dissertation. Rutgers University.
- McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis*, 9(2), 171-178.
- Paulsell, D., Austin, A. M. B., & Lokteff, M. (2013). *Measuring implementation of early childhood interventions at multiple systems levels* (OPRE Research Brief 2013-16). Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Piasta, S. B., Justice, L. M., McGinty, A., Mashburn, A., & Slocum, L. (2015). A comprehensive examination of preschool teachers' implementation fidelity when using a supplemental language and literacy curriculum. *Child Youth Care Forum*, 44(5): 731-755.
- Ryan, S. (2004). Message in a model. *Educational Policy*, 18(5), 661-685.
- Ryan, S., Hornbeck, A., & Frede, E. (2004). Mentoring for change: A time use study of teacher consultants in preschool reform. *Early Childhood Research and Practice*, 6(1), <https://ecrp.illinois.edu/v6n1/ryan.html>.
- Tobin, J., Hsueh, Y., & Karasawa, M. (2009). *Preschool in three cultures revisited: China, Japan, and the United States*. Chicago, IL: Chicago University Press.
- Tobin, J., Wu, Y. H. D., & Davidson, D. H. (1989). *Preschool in three cultures: Japan, China, and the United States*. Chicago, IL: Chicago University Press.
- Vavrus, S., & Bartlett, L. (2006). Comparatively knowing: Making a case for the vertical case study. *Current Issues in Comparative Education* 8(2), 95-103.
- Weiland, C. (2018). Commentary: Pivoting to the "how": Moving preschool policy, practice, and research forward. *Early Childhood Research Quarterly*, 45, 188-192, <https://doi.org/10.1016/j.jecresq.2018.02.017>.
- Whitebook, M., Phillips, D., & Howes, C. (2014). *Worthy work, still unlivable wages: The early childhood workforce 25 years after the National Child Care Staffing Study*. Berkeley, CA: Center for the Study of Child Care Employment, University of California.
- Whitebook, M., & Ryan, S. (2011). *Degrees in context: Asking the right questions about preparing skilled and effective teachers of young children*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.
- Whitebook, M., Ryan, S., Kipnis, F., & Sakai, L. (2008). *Partnering for preschool: A study of center directors in New Jersey's mixed delivery Abbott program*. Berkeley, CA, University of California, Center for the Study of Child Care Employment.
- Wilinski, B. (2017). *When pre-K comes to school: Policy, partnerships, and the early childhood education workforce*. New York: Teachers College Press.



SECTION 3, CHAPTER 12

**EQUITY AS A PERSPECTIVE FOR
IMPLEMENTATION RESEARCH IN
THE EARLY CHILDHOOD FIELD**

Milagros Nores, Ph.D., National Institute for Early Education Research



A data collector in a U.S. preschool classroom observed a teacher call security because she perceived a child as being disrespectful and difficult. The preschooler was observed being removed from the classroom. This occurred during a standard observation of classroom quality in one of our research projects. Standard research practices with respect to processes in early childhood may end with the classroom being given a high “negative discipline” score. Because of the limitations of standard protocols, unanswered questions remain when looking at the data. Was the child black? A boy? Hispanic? All three? To the extent that research on processes inquires more deeply into these questions, it may more fully account for how programs operate and are implemented and shed light on the biases that are reproduced in early childhood systems.

This anecdote is one of many in the research that demonstrates how the measures we use and the protocols we enact provide only a limited view of the issues and problems embedded in the implementation of policies and practices in early childhood. This chapter therefore delves into the question of equity and why equity matters in early childhood education and development (ECED) programs. It also explores the central role of research in deciphering how and when ECED programs do in fact contribute to equity (or not), and, more specifically, how equity can be embedded in evaluation designs.

Equity is “the absence of systematic and potentially remediable differences in one or more aspects ... between groups of people characterized socially, geographically, or demographically” (Starfield, 2007, p. 483). Inequities may be rooted in discrimination due to gender, disability, race/ethnicity, language, minority status, or religion; structural poverty; geographic isolation; weak governance; and cultural norms (Bamberger & Segone, 2011). Critical race theory—which contends that research and discussion of social inequity, and school inequity in particular, should consider race and racism—has been central to strengthening the ECED field’s conceptualization of inequities (Ladson-Billings, 2004).

A vision of increasing equity inspired the growth of ECED programs that reduce disparities, readiness gaps, and inequities at the starting gate, and equalizing the playing field at kindergarten entry—goals that are part of the mission of many preschool programs across the country.¹ This vision and mission derive from years of research on how preschool programs may affect not only middle-class children but also disadvantaged, special needs, and dual language children, among others (Yoshikawa, et. al, 2013).

¹ For example, Head Start states “that every child, regardless of circumstances at birth, has the ability to succeed in life” (<https://www.nhsa.org/about-us/mission-vision-history>). The Abbot preschool program implementation guidelines state that “intensive, high-quality preschool programs can close much of the early achievement gap for lower income children” (<https://www.nj.gov/education/ece/guide/impguidelines.pdf>). The Seattle preschool program includes a “commitment to early learning as the foundation for future academic success and a strategy for closing opportunity gaps” (<https://www.seattleschools.org/cms/One.aspx?portalId=627&pageId=33661301>).

But not all programs are created equal (Yoshikawa et al., 2013; Camilli, Vargas, Ryan, & Barnett, 2010). Research on program quality and processes and on implementation has helped us understand why some programs work and some do not, and why some work for some children and not others—information that is crucial to an equity-based evaluation (Bamberger & Segone, 2011). Research can not only help bring to light what works in the early years but can also document how programs contribute to increasing equity (or reducing inequity) and at what point in the education process they do so. That is, it can help us understand the effectiveness, efficiency, relevance, impact, and sustainability of ECED programs with respect to equity goals.

However, research on what occurs in preschools classrooms, teacher practices, interactions, the effectiveness of programs or preschool curricula, and ultimately, their effect on children cannot be separated from the biases and inequities that children and families may experience in the education process and the social structures in which schools and individuals are embedded. Biases and racism are present as early as preschool and kindergarten, whether it be in teachers' perceptions of Black children's behavior (Ladson-Billings, 2011; Yates & Marcelo, 2014), in perceptions of Black girls as less innocent and more adult-like, a perception known as adultification (Epstein, Blake, & Gonzalez, 2017), or in children's own perceptions of race (Farago, Sanders, & Gaias, 2015). More recently, research on preschool expulsion has also shown how implicit biases in preschool may also be determining disciplinary behavior early on (Mitchell, Fonseca, & LaFave, 2016).² To the extent that we care about equity, research should, when feasible, measure the degree to which processes and programs in early childhood reduce or exacerbate inequities and what exactly in the program's design or its implementation is contributing to these results.

Yet we cannot escape the fact that research itself—and the measures, researchers, observers, interviewers and other agents of research—may introduce biases of its own to any evaluation process. And if questions pertaining to equity are not asked, then equity is not assessed at all.

All of this matters in terms of research validity (American Evaluation Association, 2011; Kirkhart, 2010, 2013). Kirkhart defines multicultural validity as the “accuracy or trustworthiness of understandings and judgments, actions, and consequences, across multiple, intersecting dimensions of cultural diversity” (2010, p. 401). She argues that validity is enhanced when attention to cultural diversity and reflection on cultural biases helps guide the choices of epistemologies, methods, and procedures. She further argues (2005) that validity is threatened when culture is ignored or diversity stereotyped.

We cannot escape the fact that research itself—and the measures, researchers, observers, interviewers and other agents of research—may introduce biases of its own to any evaluation process.

² Research on implicit biases and behavior expectations of teachers reveals that preschool teachers are more likely to expect challenging behaviors from black children and, in particular, black boys (Gilliam, Maupin, Reyes, Accavitti, & Shic, 2016). The authors define “implicit bias” as the “automatic and unconscious stereotypes that drive people to behave and make decisions in certain ways” (p. 3).

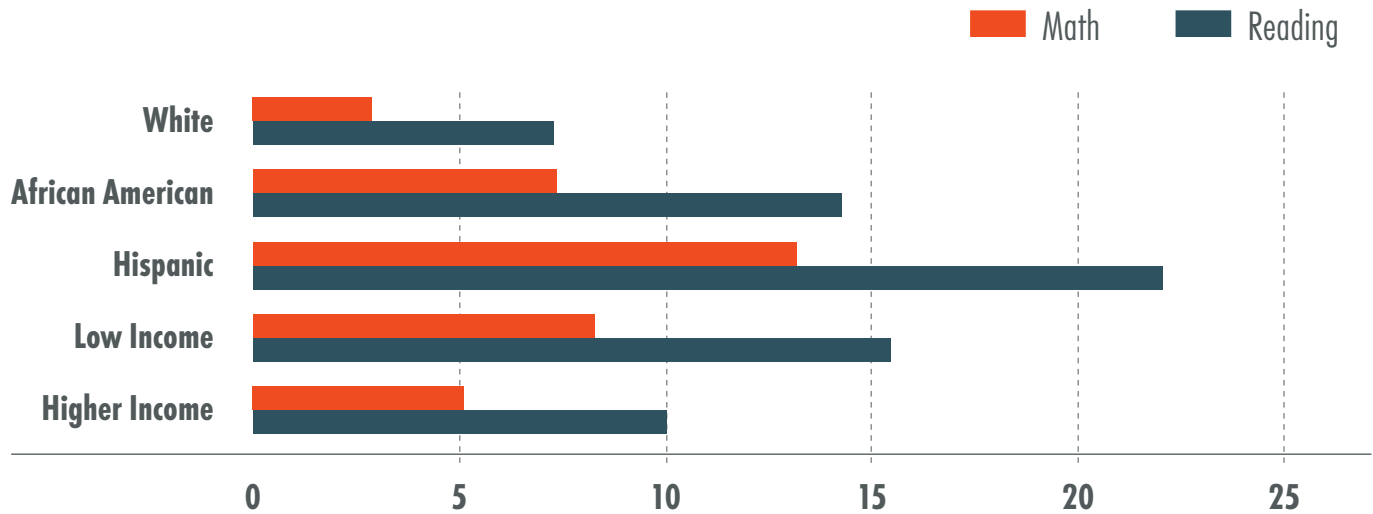
Equity in research implies capturing the extent to which programs, policies, and interventions reduce or increase inequities, validly defining inequities in relation to the context and the disadvantages that are present, and integrating the concept of equity into all components of research, from the questions asked to the analysis and interpretation stage. In sum, understanding equity means being able to answer questions that attend to equity concerns. Who are the less advantaged, and how does this evaluation capture their experience with ECED policies and programs?

EARLY CHILDHOOD PROGRAM EFFECTS AND EQUITY

Research on early childhood has provided quite robust evidence regarding the importance of preschool and has still more to contribute in terms of structure, curriculum, program features, and leadership, among other aspects (Bowne et al., 2017). Research on quality preschool programs has shown that small- and large-scale public programs can have long-term and substantial effects on children’s developmental trajectories (Camilli et al., 2010, McCoy et al., 2017, Yoshikawa et al., 2013). Research also shows that while all children can benefit significantly, children from low-income backgrounds (Gormley, Gayer, & Phillips, 2008; Weiland & Yoshikawa, 2013), children with special needs (Phillips & Meloy, 2012; USHHS, 2010; Weiland, 2016), dual language children (Barnett et al., 2007; Bloom & Weiland, 2015; Bumgarner & Brooks-Gunn, 2015; Dickinson & Porche, 2011; Goldenberg, 2012; Puma et al., 2010; Slavin, Madden, Calderón, Chamberlain, & Hennessy, 2011; Wilson Dickinson, & Rowe, 2013), and children from a racial or ethnic minority background (Gormley, Gayer, & Phillips, 2008; Weiland & Yoshikawa, 2013) may benefit as much or more than others.

For example, studies of universal preschool programs in Boston (Weiland & Yoshikawa, 2013) and Tulsa (Gormley, Gayer, & Phillips, 2008) have found positive effects on children’s math and reading achievement scores (among others) at kindergarten entry. These effects were larger for low-income, African American, and Hispanic children. Figure 1 (based on Friedman-Krauss, Barnett, & Nores, 2016, p. 11) shows average effects across these two programs reported in months of learning. Using these averages, Friedman-Krauss et al. (2016) have estimated that on average, universal programs of the same quality could reduce gaps in math skills for African Americans by 45% and for Hispanics by 78% and eliminate reading gaps for both these groups of children. While individual state population compositions and readiness gaps differ, with some of them exhibiting large percentages of white low-income or native low-income children, these projections have nationwide implications. A meta-analysis that covers 23 early education programs from the perspective of gender equity (Magnuson et al., 2016) finds that effects are generally similar for boys and girls. Differences are observed mostly across middle childhood, when the programs seem to have a greater impact on boys with respect to grade retention and special education placement.

Figure 1. Average positive effects across two universal preschool programs in months of learning.



Notes: The low-income category includes children with household incomes at or below 200% of the federal poverty guidelines, while the higher-income category includes children with household incomes above 200% of the federal poverty guidelines. Source: Friedman-Krauss et al., 2016, p. 11.

HOW DO WE DEFINE EQUITY IN RESEARCH?

Equity-focused implementation research can be understood as “analyzing the impact of internal and external processes, as well as foundational assumptions and interpersonal engagement, on marginalized and underserved individuals and communities” (Spark Policy Institute, 2014) within the process of implementation research, that is, within the process of inquiring how programs, policies, and individual practices are enacted in real-world settings (Halle, 2020). Equity, therefore, is a perspective a researcher brings to the research process that calls for understanding the “complexity and multidimensionality of context, culture and power as fundamental elements to be addressed in evaluation” (Dean-Coffey, Casey, & Caldwell, 2014, p. 84). Ultimately, the goal of equity in research is to ensure that research components capture whether a program is working toward reducing inequities and is validly defining these inequities in relation to the context and populations at hand and that evaluations of processes and programs are not introducing biases that reduce the chances of understanding whether the program works and, if it does, for whom.

A similar and highly interconnected concept (or evaluation paradigm) that has gained traction as a mechanism with an equity perspective is cultural competence, which involves understanding the unique and defining characteristics of different populations with which researchers engage (Harvard Clinical and Translational Science Center, 2010). The culturally competent researcher values diversity, understands the dynamics of the differences among subpopulations,

and has the capacity to adapt to diversity (Shiu-Thornton, 2003).³ An analogous concept is cultural responsiveness, which is defined as “a theoretical, conceptual and inherently political position that includes the centrality of, and attunement to, culture in the theory and practice of evaluation” (Hood, Hopson, & Kirkhart, 2015, p. 283).

Lastly, intersectional approaches “challenge practices that isolate and prioritize a single social position and emphasize the potential of varied inter-relationships of social identities and interacting social processes in the production of inequities” (Bécares & Priest, 2015, p. 3). From a research perspective, intersectionality means adopting an approach to the subject of study in which multiple marginalizations (by sex, gender, race, ethnicity, income, social class, education, age, sexuality, immigration history, geography, among others) are considered, rather just a single difference. Bauer (2014) proposes that these should be considered in an additive scale (in quantitative studies, this relates to measuring the combined added effect of two characteristics as different from the sum of each individual characteristic alone). Such approaches can further the field’s capacity to specifically document inequities in early childhood within intersectional groups—African American boys, African American girls, Native American girls, Hispanic immigrant children, or Muslim immigrant children, for example (Ford & Harawa, 2010). As Bauer (2014) points out, carefully considering intersectional issues can reduce measurement bias, improve construct validity, allow identification of heterogeneity of effects, and avoid the problem of average total effects that do not represent any true group (see also Whitesell, 2017).

Equity, cultural competence and responsiveness, and intersectional approaches all interconnect in central ways in the design, collection, analyses, and interpretation stages of the research work. At their core is an emphasis on understanding the complexity of social and power dynamics and an explicit attempt to recognize, measure, and assess differences, as well as reduce biases (as much as possible) and employ culturally appropriate methods. In essence, as we assess early education programs, we must take into account that these programs take place in various settings and contexts; that they have differential effects on children of different racial, ethnic, language backgrounds, of differing genders, and with differing needs (among other aspects); that children in different types of settings (e.g., urban versus rural) may have different levels of cumulative deprivation; and that all of this is central to understanding (and measuring) differences, effectiveness of processes, interactions, curriculums, and detractors and contributors throughout. At the same time, researchers should minimize any biases introduced by the research itself and strive to comprehend any cultural limitations to its methods, instruments, collection processes, or analyses.

These processes are applicable regardless of the type of research. The discussion in the next section recognizes that this may encompass (but not be limited to) basic science research, clinical or randomized trial research, ethnographic research, mixed methods research, or community-based participatory research, among others.

³ In addition, the American Evaluation Association defines it as a process of learning and relearning, awareness of self and one’s cultural position, refraining from assuming a full understanding of stakeholder perspectives, and recognizing dynamics of power (2011, p. 3).

The same is true for process, progress, and summative evaluations. Process evaluations focus in particular on how program or project components interconnect and are being implemented. Equity in this sphere would ensure not only that implementation is being documented but that the methods and measures used for the process apply an equity lens for interpreting progress (Frierson Hood, Hughes, & Thomas, 2010). Progress evaluation focuses on whether progression toward stated goals is taking place. Equity questions that may be put forward include whether the goals respond to different types of individuals and needs and whether there is any indication of equitable progress. Summative evaluations are intended to show a program's effectiveness. The role of equity here is to assess whether gains are inclusive and to situate the results in the contexts and environments necessary to interpret them adequately.

WHY ARE THESE APPROACHES IMPORTANT FOR RESEARCH?

Grounding research in equity-based perspectives, cultural competence, and intersectional approaches enhances it in various ways. Cultural competence heightens effective interactions between researchers and participants in both qualitative and quantitative research. This happens because researchers actively seek to engage with the diverse perspectives and segments of the community, respect the cultures represented, and remain aware of how their own backgrounds and experiences limit or enhance the conduct of research (American Evaluation Association, 2011).

Cultural knowledge comes from understanding differences, similarities, and inequities that may be structurally determined.

More specifically, Papadopoulos and Lees (2001) put forward a model for the development of culturally competent researchers based on cultural awareness, cultural knowledge, and cultural sensitivity. These authors developed their framework in the nursing field, but these concepts can be incorporated into the more general notion of cultural competence. They illuminate how cultural competence can enhance interactions between researchers and participants via awareness, which they define as a process in which researchers reflect

“on how their own values, perceptions, behavior, or presence and those of respondents can affect the data they collect” (p. 260). Cultural knowledge comes from understanding differences, similarities, and inequities that may be structurally determined. Cultural sensitivity derives from a true partnership with the agents of research. The authors argue that matching ethnicities of interviewers and participants, for example, encourages the latter, although it does not guarantee it (Frierson, Hood, & Hughes, 2002). Researchers, they add, should also ensure that all research components, including design, data collection, analyses, interpretation, and dissemination, are guided by cultural awareness, knowledge, and sensitivity. Cultural competence is not a series of steps that a researcher carries out apart from an evaluation or research process; rather, it undergirds how that process is carried out (Frierson et al., 2010; Hood, Hopson, & Kirkhart, 2015).

Integrating these concepts into various research components can ensure that racism is challenged, ethnocentricity is considered, and essentialism (blaming culture for results observed for a group) is avoided.

Another strength of foregrounding equity and cultural responsiveness is that it improves communication with racial and ethnic minorities or other groups (for example, language minorities) in research studies. It also produces a more accurate representation of cultural processes and practices because the researcher understands and effectively responds to factors that might influence individuals' participation, whether they be children, families or staff members, such as their history, their circumstances, and current policies that affect them. Kien Lee (2007) provides examples: an evaluator in a Native American community will be much better equipped if she understands the history of oppression, sovereignty struggles, and research misrepresentation that Native Americans have faced (see also LaFrance & Nichols, 2010). Likewise, evaluators working with women need to understand and account for existing gender roles. Similarly, working in settings with large immigrant populations requires understanding immigration policy (see, for example, Allman & Slavin, 2018).

An equity lens also incorporates an adequate representation of groups (Hood, Hopson, & Kirkhart, 2015). This requires purposeful methods for securing consent, sampling, and recruiting. Intersectional or multicultural representations across categories (race, ethnicity, religion, gender, age, language, disability, and socioeconomic background) allow for an understanding of differences and inequities as well as of pathways for inequities (Kirkhart, 2010; Bécares & Priest, 2015). The categorical labels that are most frequently used to represent individual characteristics (race, ethnicity, gender, age, language, or disability) do not capture the whole of human diversity because diversity is also constituted within categories, and it is crucial to understand the intersecting cultural identifications that these combinations represent (Kirkhart, 2010).

When it comes to measuring implementation in ECED programs, Aboud and Prado (2018) suggest that there may be various alternatives depending on the goal of implementation, whether it is piloting a program to determine feasibility or examining a well-developed program, in which the focus would likely be on quality and fidelity, among others. They explain that most ECED programs can be categorized as being delivered to children either directly (e.g., preschool) or indirectly via caregivers (e.g., home visiting). In this context, equity will come into play through the effects of the program on children (e.g., when assessing a pilot), the practices and processes observed by caregivers and teachers, curriculum enactment, enrollment practices, exclusion/inclusion of children/parents, attendance rates of children/home visitors, or expulsion practices, among other things.

COMPONENTS OF RESEARCH

Thomas and McKie (2006) provide examples of how researchers' values, beliefs, and biases can compromise an evaluation process. The questions asked and the questions not asked, what is focused on versus what is minimized, the evaluation approach selected versus the one discarded, the data collected versus the data disregarded, the interpretations made, and how and to whom the results are presented can all undermine an evaluation.

An approach to research that truly incorporates equity requires integrating equity concepts across all these components, from questions asked to interpretation (Hood, Hopson, & Kirkhart, 2015).

THEORETICAL FRAMEWORK AND EVALUATION QUESTIONS

Research and evaluation are grounded in theory: evaluation theories, social science theories, program theories, and theories of change, all of which signify implicit and explicit assumptions about how programs or practices operate and how individuals respond to such programs or practices (American Evaluation Association, 2011). Therefore, as the theoretical framework for research is developed, researchers should explicitly examine the values, beliefs, and approaches embedded in it as well as whether it fits the "evaluated" population. The American Evaluation Association (2011) advocates that researchers thoughtfully consider alternative competing frameworks, assess fit of theory to the context, and pay attention to complex power explanations within systems.

A crucial step in any evaluation is defining the questions to be addressed. The questions and how they are worded are critical to setting the evaluation on the right path. They may address needs and strengths, processes, use of resources, progress toward outcomes, and effectiveness, among other things (Hood, Hopson, & Kirkhart, 2015). Thinking in terms of equity when developing research questions entails considering whether processes are strengthened or hindered by culture, which may point to cultural fitness, on the one hand, or suggest that adaptations are needed, on the other. It also requires understanding the distribution of benefits. For example, is the program benefiting some groups more than others? Is the program reducing initial disparities among individuals? Are research questions addressing differences across and within relevant groups? Are processes reducing inequities? And if the answer to any of these questions is yes, implementation researchers must explain why. For example, are any subgroups with lower rates of absenteeism, and if so, why? Does any group show high teacher turnover and, if so, which teachers and why?

DESIGN AND SAMPLING

Design encompasses the sources and type of data, the individuals from whom evidence will be drawn, the approach (quasi-experimental, experimental, ethnographic, case study, or mixed methods), and the timing, among other aspects. Here equity will define who is represented, whether differences between and within groups can be

assessed, and how much information is collected in processes that will contextualize and identify the sources of differences across groups. Examples of questions researchers can use to guide design and sampling are who is included with this design, who is excluded, and whether the different groups that make up the target population will be well represented.

The degree to which design decisions bear on who is included and who is excluded is a central equity consideration. In quantitative designs, researchers pay close attention to selection bias and its implication for the design, the analytical strategy, and the interpretation. Heckman (1990) defines selection bias as the “distorted representation of a true population as a consequence of a [nonrandom] sampling rule” (p. 201). Distorted selection rules are likely the outcome of self-selection decisions by families, children, teachers, principals, and so forth. And selection rules introduced by the design may also generate selection biases. For example, say we are studying a program that assesses the impact of a specific racial justice curriculum, but only parents who are interested opt in to these classes, while parents of other children just continue in general education classes. The evaluation will then confound program effects with the effects of families or home environments. These parents are particularly motivated by this type of content, which very likely impacts other choices and behaviors in the home and, ultimately, would also impact the outcome of interest. If we understand the selection rules that define who is the target of a program or the intervention focus of a particular study, we can understand who is left at the margins, whether the design can find ways to include them, and to what degree the research is valid and generalizable (Willis & Rosen, 1979; Grimes & Shulz, 2002). Randomization helps to avoid selection bias and create comparable groups at baseline, yet it does not eliminate biases, such as those due to measurement, attrition, or low response rates, from other evaluation components (Torgerson & Torgerson, 2003).

Closely tied to the issue of selection bias are process aspects such as barriers to participation in the intervention or program evaluated, as well as in the study itself. It is important to create design and research strategies that address participation and take into account timing and sampling. Will the researchers be able to distinguish differences across disadvantaged groups with the design and sample size that is proposed? That is, is the statistical power sufficient for quantitative inquiries such as subgroup analyses, and are all groups in fact represented so that the investigation is qualitatively adequate? Sampling also has key implications for coherence and biases in qualitative methods, where researchers need to specify what is included or excluded when it comes to sample size, sampling strategy (random sampling, convenience sampling, stratified sampling, cell sampling, quota sampling, or a single-case selection strategy), and sample source (Robinson, 2014). For example, in both quantitative and qualitative studies, we pay close attention to teachers but rarely include teacher assistants as informants on quality. Yet they often more closely represent the children’s culture than do the lead teachers (Figueras-Daniel, 2016). Similarly, the literature often does not follow up on what drives program attrition, and attendance issues and costs of ECED programs are rarely reported (Connolly & Olson, 2012; Logan, Piasta, Justice, Schatschneider, & Petrill, 2011; Greenwood et al., 2018).

INSTRUMENTS

Instruments may themselves introduce biases. The American Evaluation Association (2011) recommends choosing data collection instruments that have been used with the populations of interest and that have shown sensitivity to those populations. This does not guarantee a lack of bias, as there is no perfect instrument. But it does make it more likely that an instrument will effectively capture increases in equities (changes over time and between groups) in the disadvantaged populations of interest. When using standardized instruments, researchers may have to review their weaknesses for particular subgroups in the population of interest. Who does the instrument not measure well? That is, researchers should reflect critically on “what constitutes meaningful, reliable, and valid data” (American Evaluation Association, 2011, p. 9), starting at the planning stage and continuing throughout data collection.

As an example, quantitative evaluations measuring the impact of specific preschool-age interventions and/or preschool programs have many times relied on the Peabody Picture Vocabulary Test (PPVT) (Dunn & Dunn, 1997; Dunn, Dunn, & Dunn, 2007). The PPVT has shown sensitivity in gauging growth in receptive English vocabulary in children identified as African American (e.g., Weiland & Yoshikawa, 2013), Hispanic (e.g., Bloom & Weiland, 2015; Weiland & Yoshikawa, 2013), and dual language learners (e.g., Bloom & Weiland, 2015; Durán, Roseth, & Hoffman, 2010; Slavin et al., 2011) across many evaluations. Despite having shown sensitivity to specific population groups, instruments may have biases that are yet unclear, and the PPVT has been challenged on the basis of limitations in assessing dual language competencies in the early years (Bandel, Atkins-Burnett, Castro, Wulsin, & Putnam, 2012). Further research could help establish measurement invariance for different subgroups. For example, Nores and Barnett (2018) have established that the PPVT-III performs equally well between English and Spanish home language speakers and between boys and girls. Because they lacked a sample with a language difference for the PPVT-IV, the authors could only replicate this process for gender difference, establishing partial measurement invariance between boys and girls for the measure. Similar analyses are needed for most measures used with preschool children and infants.

Including individuals from the population of interest in the processing of vetting instruments that are being piloted would help reduce biases (O’Brien et al., 2006; O’Brien, Harris, Beckman, Reed, & Cook, 2014). This vetting process could take culture, race, ethnicity, and language into account as well (O’Brien et al., 2006; Public Policy Associates, 2015; see Appendix). The same is true when translating or adapting an instrument (Dettlaff & Fong, 2011).

We also have much more to learn about the weak associations between existing measures of classroom quality and children’s learning (Burchinal, 2018). Researchers have started to push for more depth or further content specialization in the process measures used in early childhood education to understand quality (Atkins-Burnett, Sprachman, Lopez, Caspe, & Fallin, 2011; Goodson, Layzer, Smith, & Rimdzius, 2004; Zaslow et al., 2016) and to measure program impact on different subgroups of children, such as dual language learners or children with special

needs (Castro, Espinosa, & Páez, 2011; Halle, Vick Whittaker, & Anderson, 2010; Peisner-Feinberg et al., 2014; Soukakou, Winton, West, Sideris, & Rucker, 2014).

Similarly, measures are starting to be developed to further inquire into leadership and climate (e.g., Pacchiano, Klein, & Hawley, 2016; Whitebook & Ryan, 2012) in early childhood education settings. These are still new in the ECED field, and pending further inquiry we do not yet know whether these measures respond to the different types of programs and different populations served.

FIELDWORK

Fieldwork encompasses ethics approvals, recruitment strategies and training of field personnel, management of data collection, consenting procedures, survey and interview protocols and procedures, focus group protocols and procedures, retention policies and strategies, and translation and interpretation services. A lot of culturally responsive work should occur at the fieldwork stage, where one-on-one interactions take place between a research team and partners in the field who are willing to be research subjects and agents.

Cultural competency assessments and frameworks are highly relevant to this stage of work. The Appendix lists questions associated with various frameworks and self-assessments regarding whether assessors require culturally competent training, how to determine criteria for choosing interviewers, and how to create a flexible process that accounts for the needs of individuals or contexts (O'Brien et al., 2006, Public Policy Associates 2015; Whitesell, 2017).

Consent strategies and issues of representation are central to any evaluation. It's critical to use strategies that promote comprehensive participation, including making accommodations for language as necessary (American Evaluation Association, 2011), and to reduce barriers to the participation of groups in the study. This is especially important because active consent already reduces representation of disadvantaged populations in education research (Bergstrom et al., 2009; Flay & Collins, 2005). Accommodations should also extend beyond the consent period, to communication, assessment, survey, interview, and all evaluation activities (American Evaluation Association, 2011); this may necessitate translation or interpretation services.

Retention policies and strategies (including incentives) should reflect the culture and the individuals or children who take part in the study. They should also be effective at reducing the impact of differential attrition of particular subgroups. This will help retain validity and preserve the capacity of the study to answer questions on equity. Research on factors affecting survey response (Edwards et al., 2002; Fan & Yan, 2010), as well as on effective retention strategies for samples (Robinson, Dennison, Wayman, Pronovost, & Needham, 2007) has shown that accounting for these factors—and for demographic differences among leadership, staff, and children—can increase response rates and reduce differential attrition.

METHODS AND ANALYSES

Initial checks at this stage should ensure that attrition and/or survey response has not been differential. That is, the processes used for design and sampling, instruments, and fieldwork should not result in a sample that is more representative of a particular category (by language, race, ethnicity, gender, immigration status, or other identification) than the target population. Did only some teachers answer the surveys? Who attended the focus groups? Who finished the assessments? Who attended the program? The training? Differences between the target group and the final sample need to be clearly reported, both because they may bias results and because they are necessary to interpret analyses.

Central equity questions at this stage include the following. Are there outcomes differences, intended and unintended? Are the data disaggregated along demographic lines so that it is possible to understand programs along lines of race, culture, socioeconomic status, language, and so forth? Were there factors that contributed to disparities (or reduced disparities)? Were there any unintended changes or consequences due to cultural/racial/ethnic considerations? (O'Brien et al., 2006, Public Policy Associates, 2015; see Appendix). The study has to have the statistical power to answer such questions across subgroups or intersections.

INTERPRETATION AND DISSEMINATION

Dissemination and interpretation should be based on all the concepts presented so far. Questions that can be addressed at this stage include the following. Are the main results consistent for all subgroups, or is there evidence of heterogeneous subgroup differences? Are interpretations of subgroup differences contextualized? Are institutional or programmatic factors that contributed to subgroup effects shown? Does the program reduce equity for participants along particular dimensions? Is it neutral? Negative? What factors are contributing to or hindering equity?

Interpretation should reflect the context studied and address whether the feedback based on race, ethnicity, gender, language, or another individual characteristic allows the program and agents of change to engage the system in long-term equitable change (O'Brien et al., 2006). As the Tribal Evaluation Workgroup (2013) puts it, "Evaluation should inform practice, program, and system improvement, providing information to answer questions that local program directors and staff have about how to better serve the children and families in their communities" (p. 23). In addition, assessing social (economic, sociological, political, and cultural) explanations of processes and outcomes, as well as the social institutions and processes that influence the generation and allocation of resources, can further support a comprehensive equity-focused agenda (Östlin et al., 2011).

Efforts such as the CONSORT, STROBE, COREQ, SRQR and SAGER guidelines have strengthened the research field by requiring consistency in reporting on quantitative and qualitative research (Schulz, Altman, & Moher, 2010; Bastuji-Garin et al., 2013; Tong, Sainsbury, & Craig, 2007; O'Brien et al., 2014; Heidar, Babor, De Castro, Tart, & Curno, 2016). Yet most of these do not address equity per se. SAGER focuses on sex and gender in reporting, COREQ addresses possible biases in qualitative designs, and more recently, the CARE guidelines (Yousafzai et al., 2018) have put forward a framework for reporting on implementation research. But even though these guidelines do not directly address equity, they require contextualizing results and thus provide an initial step toward strengthening reporting in implementation studies.

CONCLUSION

In essence, addressing equity in research implies capturing the extent to which programs, policies, and interventions reduce or increase inequities, validly defining inequities in relation to the context and the disadvantages that participants in programs face, and taking care that the research process itself does not introduce biases. All of this is of central importance in the context of current ECED policies that aim to reduce inequities and disadvantages before kindergarten entry.

Addressing equity in this context includes (although is not limited to) going beyond a consideration of individual race, gender, or ethnic associations that is currently the more common approach in the field. Research needs to further examine intersections among different social hierarchies and identities; explore cumulative impacts, levels, pathways, and social (economic, sociological, political, and cultural) explanations; consider the dynamic nature of inequities; and assess social institutions and processes that influence the allocation of resources and its social determinants.

In research, the concept of equity, together with cultural competence, cultural responsiveness, and intersectionality, can permeate all components and phases of research. An equity lens makes the research process more responsive to the equity goals of early childhood education, takes into account existing disadvantages, and leads to processes that make it easier to engage agents and individuals in long-term equity change. Only by understanding what's working, what is not, and why, with the intention of advancing equity across children and families, can research strongly support the development of policies for all of our children.

References

- Aboud, F. E., & Prado, E. L. (2018). Measuring the implementation of early childhood development programs. *Annals of the New York Academy of Sciences*, 1419(1), 249-263.
- Allman, K. R., & Slavin, R. E. (2018). Immigration in 2018: What is a teacher educator to do? *The Teacher Educator*, 53(3), 236-243.
- American Evaluation Association. (2011). *Public statement on cultural competence in evaluation*. Fairhaven, MA: American Evaluation Association. <https://www.eval.org/p/cm/ld/fid=92>.
- Atkins-Burnett, S., Sprachman, S., Lopez, M., Caspe, M., & Fallin, K. (2011). The Language Interaction Snapshot (LISn): A new observational measure for assessing language interactions in linguistically diverse early childhood programs. In C. Howes, J. T. Downer, & R. C. Pianta (eds.), *Dual language learners in the early childhood classroom* (pp. 117-146). Baltimore, MD: Brookes Publishing.
- Bamberger, M., & Segone, M. (2011). *How to design and manage equity-focused evaluations*. New York: UNICEF Evaluation Office.
- Bandel, E., Atkins-Burnett, S., Castro, D. C., Wulsin, C. S., & Putnam, M. (2012). *Examining the use of language and literacy assessments with young dual language learners (final report)*. Chapel Hill, NC: University of North Carolina, FPG Child Development Institute, CECER-DLL.
- Barnett, W. S., Yaroz, D.J., Thomas, J., Jung, K., & Blanco, D. (2007). Two-way immersion in preschool education: An experimental comparison. *Early Childhood Research Quarterly*, 22(3), 277-293.
- Bastuji-Garin, S., Sbidian, E., Gaudy-Marqueste, C., Ferrat, E., Roujeau, J. C., Richard, M. A., Canoui-Poutrine, F. 2013. Impact of STROBE statement publication on quality of observational study reporting: Interrupted time series versus before-after analysis. *PLoS One*, 8(8): e64733.
- Bauer, G. R. (2014). Incorporating intersectionality theory into population health research methodology: Challenges and the potential to advance health equity. *Social Science and Medicine*, 110(3), 10-17.
- Bécares, L., & Priest, N. (2015). Understanding the influence of race/ethnicity, gender, and class on inequalities in academic and non-academic outcomes among eighth-grade students: Findings from an intersectionality approach. *PLoS One*, 10(10), e0141363.
- Bergstrom, J. P., Partington, S., Murphy, M. K., Galvao, L., Fayram, E., & Cisler, R. A. (2009). Active consent in urban elementary schools: An examination of demographic differences in consent rates. *Evaluation review*, 33(5), 481-496.
- Bloom, H. S., & Weiland, C. (2015). *Quantifying variation in Head Start effects on young children's cognitive and socio-emotional skills using data from the National Head Start Impact Study*. New York: MDRC. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2594430.
- Bowne, J. B., Magnuson, K. A., Schindler, H. S., Duncan, G. J., & Yoshikawa, H. (2017). A meta-analysis of class sizes and ratios in early childhood education programs: Are thresholds of quality associated with greater impacts on cognitive, achievement, and socioemotional outcomes? *Educational Evaluation and Policy Analysis*, 39(3), 407-428.
- Bumgarner, E., & Brooks-Gunn, J. (2015). The association between early care arrangements, quality, and emergent bilingual Latino American children's math and literacy skills in English. *Early Childhood Research Quarterly*, 30(1), 32-44.
- Burchinal, M. (2018). Measuring early care and education quality. *Child Development Perspectives*, 12(1), 3-9.
- Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record*, 112(3), 579-620.

- Castro, D., Espinosa, L., & Páez, M. (2011). Defining and measuring quality in early childhood practices that promote dual language learners' development and learning. In M. Zaslow, I. Martinez-Beck, K. Tout., & T. Halle (Eds.), *Quality measurement in early childhood settings* (pp. 257-280). Baltimore, MD: Brookes Publishing.
- Connolly, F., & Olson, L. S. (2012). *Early elementary performance and attendance in Baltimore City schools' pre-kindergarten and kindergarten*. Baltimore, MD: Baltimore Education Research Consortium.
- Dean-Coffey, J., Casey, J., & Caldwell, L. D. (2014). Raising the bar—Integrating cultural competence and equity: Equitable evaluation. *The Foundation Review*, 6(2), 81-94.
- Dettlaff, A. J., & Fong, R. (2011). Conducting culturally competent evaluations of child welfare programs and practices. *Child welfare*, 90(2), 49-68.
- Dickinson, D. K., & Porche, M. V. (2011). Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development*, 82(3), 870-886.
- Dunn, L. M., & Dunn, L. L. (2007). *Peabody picture vocabulary test* (4th ed.). Minneapolis: Pearson Assessments.
- Dunn, L. M., Dunn, L. L., & Dunn, D. M. (1997). *Peabody picture and vocabulary test* (3rd ed.). Circle Pines, MN: American Guidance Services.
- Durán, L. K., Roseth, C. J., & Hoffman, P. (2010). An experimental study comparing English-only and transitional bilingual education on Spanish-speaking preschoolers' early literacy development. *Early Childhood Research Quarterly*, 25(2), 207-217.
- Edwards, P., Roberts, I., Clarke, M., DiGiuseppi, C., Pratap, S., Wentz, R., & Kwan, I. (2002). Increasing response rates to postal questionnaires: systematic review. *BMJ*, 324(7347), 1183.
- Epstein, R., Blake, J., & Gonzalez, T. (2017). *Girlhood interrupted: The erasure of black girls' childhood*. Washington, DC: Georgetown Law Center on Poverty and Inequality.
- Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in human behavior*, 26(2), 132-139.
- Farago, F., Sanders, K., & Gaias, L. (2015). Addressing race and racism in early childhood: Challenges and opportunities. In J. Sutterby (Ed.), *Discussions on Sensitive Issues* (pp. 29-66). Bingley, UK: Emerald Group Publishing Limited.
- Figueras-Daniel, A. (2016). *Key influences on the quality and outcomes of preschool education for dual language learners: Professional development and bilingual staffing patterns* (unpublished doctoral dissertation). Rutgers University-Graduate School, New Brunswick.
- Flay, B. R., & Collins, L. M. (2005). Historical review of school-based randomized trials for evaluating problem behavior prevention programs. *Annals of the American Academy of Political and Social Science*, 599(1), 115-146.
- Ford, C. L., & Harawa, N. T. (2010). A new conceptualization of ethnicity for social epidemiologic and health equity research. *Social Science and Medicine*, 71(2), 251-258.
- Friedman-Krauss, A., Barnett, W. S., & Nores, M. (2016). *How much can high-quality universal pre-k reduce achievement gaps?* Washington, DC: Center for American Progress.
- Frierson, H. T., Hood, S., & Hughes, G. (2002). Strategies that address culturally responsive evaluation. In J. Frechtling (Ed.), *The 2002 user-friendly handbook for project evaluation* (pp. 63-73). Alexandria, VA: National Science Foundation.

CHAPTER 12 EQUITY AS A PERSPECTIVE FOR IMPLEMENTATION RESEARCH IN THE EARLY CHILDHOOD FIELD

- Frierson, H. T., Hood, S., Hughes, G. B., & Thomas, V. G. (2010). A guide to conducting culturally responsive evaluations. In National Science Foundation, *The 2010 user-friendly handbook for project evaluation* (pp. 75-96). Alexandria, VA: National Science Foundation.
- Gilliam, W. S., Maupin, A. N., Reyes, C. R., Accavitti, M., & Shic, F. (2016). *Do early educators' implicit biases regarding sex and race relate to behavior expectations and recommendations of preschool expulsions and suspensions?* Research Study Brief. New Haven, CT: Yale University, Yale Child Study Center.
- Goldenberg, C. (2012). Research on English learner instruction. In M. Calderón (Ed.), *Breaking through: Effective instruction and assessment for reaching English learners* (pp. 39-61). Bloomington, IN: Solution Tree Press.
- Goodson, B. D., Layzer, C., Smith, W. C., & Rimdzius, T. (2004). *Observation measures of language and literacy instruction (OMLIT)*. Cambridge, MA: Abt Associates.
- Gormley, W., Gayer, T., & Phillips, D. A. (2008). Preschool programs can boost school readiness. *Science*, June 27, 1723-1724.
- Gormley, W., Gayer, T., Phillips, D. A., & Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology*, 41(6), 872-884.
- Greenwood, P. W., Karoly, L. A., Everingham, S. S., Houbé, J., Kilburn, M.R., Rydell, C. P, . . . & Chiesa, J. (2018). Estimating the costs and benefits of early childhood interventions: Nurse home visits and the Perry Preschool. In B. Welsh, D. P. Farrington, & L. W. Sherman (Eds.), *Costs and benefits of preventing crime* (pp. 123-148). New York: Routledge.
- Grimes, D. A., & Schulz, K. F. (2002). Bias and causal associations in observational research. *The Lancet*, 359(9302), 248-252.
- Halle, T. (2020). How implementation science and improvement science can work together to improve early care and education (this volume).
- Halle, T., Vick Whittaker, J. E., & Anderson, R. (2010). *Quality in early childhood care and education settings: A compendium of measures* (2nd ed.) Washington, DC: Child Trends.
- Harvard Clinical and Translational Science Center (2010). *Cultural competence in research*. Cambridge, MA: Program for Cultural Research, Harvard Catalyst, Program for Faculty Development and Diversity.
- Heckman, J. J. (1990). *Selection bias and self-selection*. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *Econometrics* (pp. 201-224). London: Palgrave Macmillan.
- Heidari, S., Babor, T. F., De Castro, P., Tort, S., & Curno, M. (2016). Sex and gender equity in research: Rationale for the SAGER guidelines and recommended use. *Research Integrity and Peer Review*, 1(1), 1-9.
- Hood, S., Hopson, R. K., & Kirkhart, K. E. (2015). Culturally responsive evaluation. In K. Newcomer, H. P. Hatry, & J. S. Wholey (eds.), *Handbook of practical program evaluation* (4th ed.) (pp. 281-317). New York: Wiley.
- Kirkhart, K. E. (2005). Through a cultural lens: Reflections on validity and theory in evaluation. In S. Hood, R. Hopson, & H. Frierson (Eds.), *The role of culture and cultural context: A mandate for inclusion, the discovery of truth, and understanding in evaluative theory and practice* (pp. 21-39). Greenwich, 36CT: Information Age Publishing.
- Kirkhart, K. E. (2010). Eyes on the prize: Multicultural validity and evaluation theory. *American Journal of Evaluation*, 31(3), 400-413.
- Kirkhart, K. E. (2013, April). Repositioning validity. Paper presented during a plenary panel, Perspectives on Repositioning Culture in Evaluation and Assessment, at CREA Inaugural Conference, Chicago.

- Ladson-Billings, G. (2004). Just what is critical race theory and what's it doing in a nice field like education. In G. Ladson-Billings & D. Gillborn (Eds.), *The RoutledgeFalmer reader in multicultural education* (pp. 49-67). London: RoutledgeFalmer.
- Ladson-Billings, G. (2011). Boyz to men? Teaching to restore black boys' childhood. *Race Ethnicity and Education*, 14(1), 7–15.
- LaFrance, J., & Nichols, R. (2010). Reframing evaluation: Defining an Indigenous evaluation framework. *The Canadian Journal of Program Evaluation*, 23(2), 13.
- Lee, K. (2007). *The importance of culture in evaluation: A practical guide for evaluators*. Denver: The Colorado Trust.
- Li, Z., Nores, M., and Barnett, W. S. (2018, April). Measurement invariance of the Peabody Vocabulary Test. Paper presented at the 2018 American Education Research Association Annual Meeting, New York.
- Logan, J. A., Piasta, S. B., Justice, L. M., Schatschneider, C., & Petrill, S. (2011). Children's attendance rates and quality of teacher-child interactions in at-risk preschool classrooms: Contribution to children's expressive language growth. *Child and Youth Care Forum*, 40(6), 457-477.
- Magnuson, K. A., Kelchen, R., Duncan, G. J., Schindler, H. S., Shager, H., & Yoshikawa, H. (2016). Do the effects of early childhood education programs differ by gender? A meta-analysis. *Early Childhood Research Quarterly*, 36(3), 521-536.
- McCoy, D. C., Yoshikawa, H., Ziol-Guest, K. M., Duncan, G. J., Schindler, H. S., Magnuson, K., . . . Shonkoff, J. P. (2017). Impacts of early childhood education on medium-and long-term educational outcomes. *Educational Researcher*, 46(8), 474-487.
- Mitchell, S., Fonseca, M., & LaFave, A. (2016). *Suspension and expulsion in preschool development states: Policies and practices*. State Technical Assistance Report. Preschool Development and Expansion Grant Technical Assistance. Herndon, VA: AEM.
- O'Brien B. C., Harris I. B., Beckman T. J., Reed D. A. & Cook D. A. (2014). Standards for reporting qualitative research: A synthesis of recommendations. *Academic Medicine*, 89(9): 1245-1251.
- O'Brien, R. L., Kosoko-Lasaki, O., Cook, C. T., Kissell, J., Peak, F., & Williams, E. H. (2006). Self-assessment of cultural attitudes and competence of clinical investigators to enhance recruitment and participation of minority populations in research. *Journal of the National Medical Association*, 98(5), 674-682.
- Östlin, P., Schrecker, T., Sadana, R., Bonnefoy, J., Gilson, L., Hertzman, C., . . . Muntaner, C. (2011). Priorities for research on equity and health: Towards an equity-focused health research agenda. *PLoS medicine*, 8(11), e1001115.
- Pacchiano, D., Klein, R., and Hawley, M.S. (2016). *Reimagining instructional leadership and organizational conditions for improvement: Applied research transforming early education*. Chicago: The Ounce of Prevention Fund.
- Papadopoulos, I., & Lees, S. (2002). Developing culturally competent researchers. *Journal of advanced nursing*, 37(3), 258-264.
- Peisner-Feinberg, E., Buysse, V., Fuligni, A., Burchinal, M., Espinosa, L., Halle, T., & Castro, D. C. (2014). Using early care and education quality measures with dual language learners: A review of the research. *Early Childhood Research Quarterly*, 29(4), 786-803.
- Phillips, D., & Meloy, E. (2012). High-quality school-based pre-k can boost early learning for children with special needs. *Exceptional Children*, 78(4), 471-90.
- Public Policy Associates. (2015). Considerations for conducting evaluation using a culturally responsive and racial equity lens. Lansing, MI: Public Policy Associates. <http://publicpolicy.com/wp-content/uploads/2017/04/PPA-Culturally-Responsive-Lens.pdf>.
- Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., . . . Ciarico, J. (2010). *Head Start impact study: Final report*. Washington, DC: Administration for Children and Families, Office of Planning, Research and Evaluation.

CHAPTER 12 EQUITY AS A PERSPECTIVE FOR IMPLEMENTATION RESEARCH IN THE EARLY CHILDHOOD FIELD

- Robinson, K. A., Dennison, C. R., Wayman, D. M., Pronovost, P. J., & Needham, D. M. (2007). Systematic review identifies number of strategies important for retaining study participants. *Journal of clinical epidemiology*, 60(8), 757-e1.
- Robinson, O. C. (2014). Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative research in psychology*, 11(1), 25-41.
- Schulz, K. F., Altman, D. G., & Moher, D. 2010. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *Journal of Pharmacology and Pharmacotherapeutics*, 1(2), 100-107.
- Shiu-Thornton, S. (2003). Addressing cultural competency in research: Integrating a community-based participatory research approach. *Alcoholism: Clinical and Experimental Research*, 27(8), 1361-1364.
- Slavin, R., Madden, N., Calderón, M., Chamberlain, A., & Hennessy, M. (2011). Reading and language outcomes of a multiyear randomized evaluation of transitional bilingual education. *Educational Evaluation and Policy Analysis*, 33(1), 47-58.
- Soukakou, E. P., Winton, P. J., West, T. A., Sideris, J. H., & Rucker, L. M. (2014). Measuring the quality of inclusive practices: Findings from the inclusive classroom profile pilot. *Journal of Early Intervention*, 36(3), 223-240.
- Spark Policy Institute. (2014). Tools for integrating an equity lens. <http://tools.sparkpolicy.com/introduction-taking-an-equity-lens>.
- Starfield, B. (2007). Pathways of influence on equity in health. *Social Science and Medicine*, 64(7), 1355-1362.
- Thomas, V. G., & McKie, B. K. (2006). Collecting and utilizing evaluation research for public good and on behalf of African American children. *The Journal of Negro Education*, 75(3), 341-352.
- Tong, A., Sainsbury, P., & J. Craig. 2007. Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal of Quality Health Care*, 19(6): 349-357.
- Torgerson, D. J., & Torgerson, C. J. (2003). Avoiding bias in randomised controlled trials in educational research. *British Journal of Educational Studies*, 51(1), 36-45.
- Tribal Evaluation Workgroup. 2013. *A roadmap for collaborative and effective evaluation in tribal communities*. Washington, DC: Children's Bureau, Administration for Children and Families, U.S. Department of Health and Human Services.
- Weiland, C. (2016). Impacts of the Boston prekindergarten program on the school readiness of young children with special needs. *Developmental Psychology*, 52(11), 1763-1776.
- Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, 84(6), 2112-2130.
- Whitebook, M., & Ryan, S. (2012). *Supportive environmental quality underlying adult learning (SEQUAL)*. Berkeley, CA: Center for the Study of Child Care Employment, University of California.
- Whitesell, N. R. (2017). *Evidence and equity: Challenges for research design*. OPRE Report #2017-76. Washington, DC: Administration for Children and Families, Office of Planning, Research and Evaluation.
- Willis, R. J., & Rosen, S. (1979). Education and self-selection. *Journal of Political Economy*, 87(5, Pt. 2), S7-S36.
- Wilson, S. J., Dickinson, D. K., & Rowe, D. W. (2013). Impact of an early reading first program on the language and literacy achievement of children from diverse language backgrounds. *Early Childhood Research Quarterly*, 28(3), 578-592.

Yates, T. M., & Marcelo, A. K. (2014). Through race-colored glasses: Preschoolers' pretend play and teachers' ratings of preschooler adjustment. *Early Childhood Research Quarterly, 29*(1), 1-11.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., . . . Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education*. New York: Foundation for Child Development.

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.

Zaslow, M., Anderson, R., Redd, Z., Wessel, J., Daneri, P., Green, K., . . . Martinez-Beck, I. (2016). Quality thresholds, features, and dosage in early care and education: introduction and literature review. *Monographs of the Society for Research in Child Development, 81*(2), 7-26.

Appendix: Self-assessments and considerations for research

The following includes a compilation of reflection or self-assessments drawn from various perspectives on cultural competence, cultural congruence, and cultural responsiveness that are organized by context, perspective, program, design and sampling, procedures and analyses, and dissemination. All of these perspectives inform research in different ways and support reflection about all stages of research.

Concept	Lee, 2007 Cross-Cultural Competence	Kirkhart, 2010 Cultural Congruence	Public Policy Associates, 2015 Cultural Responsiveness	O'Brien et al., 2006 Cultural Competence
Context	How do people from this culture typically greet each other?	Have you learned the history of this community and of the evaluand?	Have you researched the cultural behavior and needs of the language population? For example, accommodations for language?	Have you decided whether cultural competency training is needed?
	Whom should I greet first if I am approaching a group of people?	Have you identified the relevant geographic boundaries and characteristics of this context?	Sought clarity on demographics and other characteristics of the local community?	Do you learn about the socioeconomic status, culture, or other aspects of the priority population and accommodate differences?
	How do people from this culture tend to view someone with authority and power?	Have you identified the strengths of this context?		
	What past experiences has the community had with researchers and evaluators?	Have you paid attention to how power is distributed through formal or informal structures?		
	Who are the typical knowledge holders in this culture?			
	What contextual conditions and structural inequities exist in this context?			

Continues on next page

Concept	Lee, 2007 Cross-Cultural Competence	Kirkhart, 2010 Cultural Congruence	Public Policy Associates, 2015 Cultural Responsiveness	O'Brien et al., 2006 Cultural Competence
Perspective	<p>What social identities and groups do I belong to?</p> <p>How might these color the lens through which I view the world?</p> <p>What social identities and groups do people who don't know me think I belong to?</p> <p>Who is knowledgeable enough to help me ensure multicultural validity?</p>	<p>Have you considered? the values espoused by the funders of this evaluand?</p> <p>If the program is built on prior empirical research, have you paid attention to participated in the original body of evidence and how culture was addressed?</p>		
Program	<p>Why is the initiative of the program important?</p> <p>What potential impact, both positive and negative, can the evaluation have on the community and beyond?</p> <p>Do I know what policies, procedures, and practices may affect the program's impact?</p> <p>Do I know what policies, procedures, and practices may affect the staff's performance in the evaluation?</p>	<p>What cultural characteristics are most salient in understanding the consumers of this program? Diverse? Homogenous?</p> <p>What cultural characteristics are most salient in understanding the providers of this program? Diverse? Homogeneous?</p> <p>What are the admission criteria? How does it restrict diversity?</p>	<p>Have you sought clarity on eligibility criteria?</p>	

Continues on next page

Concept	Lee, 2007 Cross-Cultural Competence	Kirkhart, 2010 Cultural Congruence	Public Policy Associates, 2015 Cultural Responsiveness	O'Brien et al., 2006 Cultural Competence
Design, Sample	<p>Who is in my sample and what do I need to know about them?</p> <p>What is the best time for me to collect data from them?</p> <p>Who should collect the data to ensure that participants feel comfortable and safe?</p>		<p>Do you routinely involve the priority population in designing some/all evaluation steps?</p> <p>Do you take race/ethnicity into account in designing survey/instrument(s)?</p> <p>Have you considered demographic or underserved populations?</p>	<p>Do you take race/ethnicity into account when designing an instrument?</p> <p>Do you consider demographic differences between leadership, staff, and children? Community context? Underserved populations?</p>
Procedures	<p>Is the location for the interview/activity easily accessible, familiar, and comfortable for the people with whom I will meet?</p> <p>What am I assuming about each group of stakeholders in the evaluation?</p>		<p>Do you find yourself changing the way you speak, and the words you use based on verbal or nonverbal cues from your recipients?</p> <p>Have you determined criteria for identification of interviewers?</p> <p>Have you decided whether interviewers need cultural competency training?</p>	<p>Do you find yourself changing verbal and nonverbal responses (words and tones) in response to who you interview?</p> <p>Do you understand the need to adapt and be flexible in your process to the needs of individuals?</p> <p>Have you determined criteria for interviewers?</p>

Continues on next page

Concept	Lee, 2007 Cross-Cultural Competence	Kirkhart, 2010 Cultural Congruence	Public Policy Associates, 2015 Cultural Responsiveness	O'Brien et al., 2006 Cultural Competence
<p>Analyses, Dissemination</p>	<p>Can the average person not steeped in evaluation terminology understand me?</p> <p>How will the findings be used by the community members, politicians, policymakers, journalists, and special interest groups?</p> <p>Will the findings place a stigma on a certain group or give the group power to access resources and improve their situations?</p> <p>What are the self-serving purposes of the research for the sponsor and the evaluator?</p>		<p>Have you checked for outcomes and differences, intended and unintended?</p> <p>Have you determined who or what is changed/affected?</p> <p>Have you observed any unintended changes or consequences due to cultural/racial/ethnic considerations?</p> <p>Do you ensure that the program is accessible to the target population?</p> <p>Do you make recommendations that focus on equity?</p> <p>Do you make use of disaggregated data along demographic lines in order to adapt your evaluation processes to the race/culture of recipients?</p>	<p>Do you disaggregate data along demographic lines to understand programs along race, culture, socioeconomic status, and language lines?</p> <p>Do you analyze and interpret outcomes, differences, and intersections?</p> <p>Do you think about how you can use the type of feedback you receive based on racial, ethnic, or other characteristics of individuals who participate in the system to engage them in long-term equitable change?</p>