



FOUNDATION  
FOR CHILD  
DEVELOPMENT

**GETTING IT RIGHT:  
USING IMPLEMENTATION RESEARCH  
TO IMPROVE OUTCOMES IN EARLY  
CARE AND EDUCATION**

<b>Introduction: Implementation research in early care and education</b>	3
<b>Section I. What does research tell us about effectiveness and implementation of ECE programs across the birth-to-eight continuum?</b>	11
Chapter 1: What Does Research Tell Us About ECE Programs?	13
Chapter 2: What Are Reasonable Expectations for ECE Program Effectiveness?	37
Chapter 3: Using a Social Determinants of Early Learning Framework to Eliminate Educational Disparities and Opportunity Gaps.	63
<b>Section 2. What still needs to be understood?</b>	87
Chapter 4: Making Prekindergarten Classrooms Better Places for Children’s Development.	89
Chapter 5: Improving Quality and Impact Through Workforce Development and Implementation Systems.	109
Chapter 6: Addressing Equity in the ECE Classroom: Equal Access and High Quality for Dual Language Learners.	131
Chapter 7: Vignette: Building a High-Quality Program—the Boston Public Schools Experience.	153
<b>Section 3. How do we get smarter? The road forward</b>	175
Chapter 8: An Overview of Implementation Research and Frameworks in Early Care and Education Research.	177
Chapter 9: Designing Implementation Research to Guide the Scale-Up of Effective Early Care and Education Across Settings.	195
Chapter 10: How Implementation Science and Improvement Science Can Work Together to Improve Early Care and Education.	223
Chapter 11: The Contributions of Qualitative Research to Understanding Implementation of Early Childhood Policies and Programs.	259
Chapter 12: Equity as a Perspective for Implementation Research in the Early Childhood Field.	275
<b>Moving ECE implementation research forward: Reflections and insights</b>	301

## Suggested citation

Foundation for Child Development. (2020). *Getting it Right: Using Implementation Research to Improve Outcomes in Early Care and Education*. New York, NY. doi: [https://www.fcd-us.org/assets/2020/06/GettingitRight\\_UsingImplementationResearchtoImproveOutcomesinECE\\_2020.pdf](https://www.fcd-us.org/assets/2020/06/GettingitRight_UsingImplementationResearchtoImproveOutcomesinECE_2020.pdf)



**INTRODUCTION**

**IMPLEMENTATION  
RESEARCH IN EARLY CARE  
AND EDUCATION**

---

Jacqueline Jones, Ph.D., Foundation for Child Development



The past 50 years have brought exceptional gains in federal, state, and local funding for early care and education in the United States. In turn, the field is working hard to make good on the evidence-based promise that quality early childhood education (ECE) can create better child and adult outcomes, particularly for underserved children. In the long run, however, if the field cannot answer implementation scale-up questions related to the specifics of how and when ECE is effective, continued support and increased investment for ECE is potentially at risk.

As the number of publicly funded ECE programs increases, policymakers will need empirical evidence to justify the taxpayer investment. Such justification will require a stronger understanding of the essential components of an ECE program's design, as well as solid evidence on which components, or constellations of components, are most effective in achieving strong outcomes for specific subgroups of children. Expectations for child outcomes must be based on the realities of the program components, the target populations, and the financial and human resources that support program implementation. We need more robust quantitative and qualitative data to ensure stronger outcomes for all young children and significantly narrow the opportunity and achievement gap for minoritized children and those living in poverty. Believing in magic will not produce strong outcomes (Brooks-Gunn, 2003). Overpromising and underdelivering will have catastrophic results for the children and families who might benefit most from ECE initiatives.

**We need more robust quantitative and qualitative data to ensure stronger outcomes for all young children and significantly narrow the opportunity and achievement gap for minoritized children and those living in poverty.**

Our standard strategy for assessing program effectiveness has been the randomized controlled trial (RCT). Such trials randomly assign some children to a group that receives a defined treatment and others to a group that does not. Assuming that all things are equal, post-treatment differences between the two groups can be attributed to the treatment's impact. This methodology, which we believe allows us to make causal inferences, provided the early evidence of ECE programs' potential in the landmark Perry Preschool and Abecedarian studies (Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010; Schweinhart et al., 2005; Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Ramey et al., 2000), and it is still considered to be the gold standard.

But this volume asserts that RCTs could be greatly enhanced by the findings from rigorous empirical data that provide contextual information about the participants, the settings, and the overall conditions under which the treatment is conducted. Throughout this volume, this type of analysis is referred to broadly as implementation research. However, our intent is not to provide a single definition of implementation research. Rather, we hope to initiate a conversation that is centered on what else needs to be explored about how ECE programs are operationalized and what shape the research might take. We hope the range of perspectives about implementation research that our chapter authors bring will serve to enrich the discussion.

As with research design, ECE programs do not follow a single model, and mean comparisons between control and treatment groups may not capture important nuances of variation in program delivery, educator skill, dosage, and so forth. For example, an RCT of the Tennessee Voluntary Prekindergarten (TN-VPK) program conducted by researchers from Vanderbilt University (Lipsey, Farran, & Hofer, 2015; Lipsey, Farran, & Durkin, 2018) highlighted that fully understanding program implementation and evaluation is a complex task.

The TN-VPK, a full-day prekindergarten program for 4-year-old children who will enter kindergarten the following school year, was evaluated using an RCT. At the end of preschool, TN-VPK attendees had significantly higher achievement scores than children who did not attend the program. But this advantage disappeared by the end of kindergarten. While the largest effects were seen among English learners regardless of their mothers' education status, by second grade, the average score of the TN-VPK treatment group was *lower* on most measures than the average score of the control group (Lipsey, Farran, & Durkin, 2018; Lipsey, Farran, & Hofer, 2015). Understandably, these surprising and disturbing findings elicited a range of interpretations, from claims of methodological error to suggestions that they were evidence of the ineffectiveness of all ECE programs.

Yet the critical question to be answered was "Why do these data look like this?" An RCT may not, by itself, answer crucial questions. Do these data reflect variability in the fidelity of implementation of the program's essential components? Are all children experiencing the program under the same conditions? Are specific subgroups of children demonstrating different responses to the intervention?

It is time to acknowledge that researchers, policymakers, and practitioners may not sufficiently understand how various components of ECE programs work or what their differential contribution is to a range of positive and negative outcomes for young children.

What does it really mean when we report the mean differences between control and treatment groups? Earlier evaluations, such as the Perry Preschool Project and the Abecedarian study were conducted when few ECE offerings were available. The treatment group received the intervention, and the control group received nothing. Given the significant growth in the number and type of ECE programs over the past 50 years, as Jeanne Brooks-Gunn and Sarah Lazzeroni note in this volume, there is no longer a "clean" control group. Families have many more ECE options, and children not assigned to a treatment group may be in an alternate type of ECE program, presenting a significant challenge in understanding to what type of group the treatment is being compared.

Is there a protocol that can guide our understanding of what is really happening in ECE programs? How should we report on the implementation and eventual evaluation of ECE initiatives such as center- and school-based programs, home visiting, family child care programs, and state-funded preschool?

This volume is intended to initiate a conversation among applied researchers who wish to use their methodological skills to help policymakers and practitioners design questions and get answers that can enhance the quality of life for all young children and their families. We also hope that policymakers will find some important questions to ask and answer as they begin to bring ECE programs to scale at the federal, state, and local levels.

## HOW THE VOLUME IS ORGANIZED

---

This volume is divided into three main sections that are intended to provide an overview of what we know about the effectiveness of ECE interventions, what remains to be understood, and what the path forward might entail.

In the first section, we describe the current state of understanding around the effectiveness of ECE interventions from birth to 8 years. Much has been learned about providing high-quality experiences for young children. The pioneering work of Brooks-Gunn, Margaret Burchinal, Linda Espinosa, Dale Farran, and Robert Pianta has advanced our knowledge of what it takes to offer high-quality experiences that promote stronger outcomes for young children. They have made us aware that we need to take a deeper look into the essential components of early childhood interventions and to meaningfully explore what works (or not), for whom, and under what conditions. We are extraordinarily fortunate to begin this volume with chapters in which these researchers describe the current state of knowledge in ECE. We see from their research that issues of equity in access to and quality of ECE programs continue to hover over the ECE landscape.

Burchinal and Farran tackle the thorny issue of the relationship between our current indicators of program quality and child outcomes, while Brooks-Gunn clarifies what child outcomes we should reasonably expect from ECE programs. Building on the foundation of our current knowledge, Iheoma Iruka identifies the potential root causes of documented disparities and proposes potentially mitigating practices and policies.

Section 2 covers what still needs to be understood in terms of content, practice, and outcomes. Farran explores what factors might strengthen outcomes for young children and develops the notion of constellations of classroom practices and specific content that show potential to enhance children's learning and development. Pianta and Bridget Hamre discuss research on effective elements of professional development and describe the need to scale effective professional development systems. They offer a set of research questions related to scaling professional development systems that highlight such issues as purpose, supports, intensity, duration, and effectiveness. In her chapter, Espinosa reviews effective program models, instructional practices, and the educator competencies needed to provide high-quality ECE for dual language learners.

The section ends with Jason Sachs’s powerful voice from the field as he reflects on building and scaling up the Boston Public Schools’ prekindergarten to second grade program. Sachs describes the intentional use of research to guide change and the realities encountered while conducting implementation research.

Section 3 explores how implementation research can help us understand ECE program effectiveness and makes a case for why we need new research approaches. JoAnn Hsueh, Tamara Halle, and Michelle Maier frame the measurement landscape needed to tell the full story of how ECE programs are actually implemented. They assert that strong implementation research is the key to how demonstrated, positive child outcomes from small-scale model ECE programs can be achieved in large-scale adaptations across populations and settings.

Maier and Hsueh provide definitions of implementation research, looking both inward at the program itself and outward at the significant organizational and contextual factors. Next, Halle’s chapter outlines distinctions among implementation science, improvement science, and program evaluation.

Sharon Ryan outlines the importance of qualitative perspectives in research design and asserts that research should continue to explore the impact of inequities that exist across the ECE workforce in terms of compensation, work environments, and benefits, especially as these relate to teacher well-being, turnover, and retention.

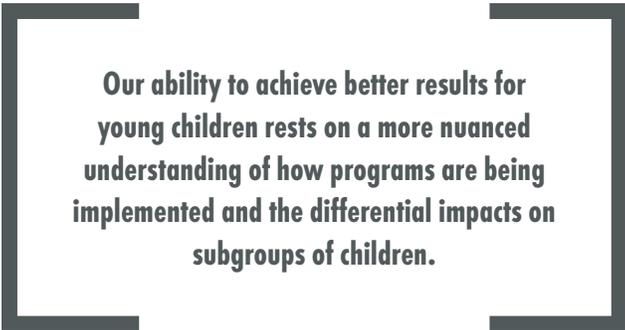
The section ends with Milagros Nores’s discussion of the need to address equity issues in research design, measures, and methodology, as well as the role of implementation research in understanding what might reduce or increase inequity.

Finally, in an afterword, Sara Vecchiotti reflects on the volume as a whole and its implications for those conducting policy-relevant implementation research.

## THE PROMISE OF IMPLEMENTATION RESEARCH

---

Our ability to achieve better results for young children rests on a more nuanced understanding of how programs are being implemented and the differential impacts on subgroups of children. Implementation research is an intriguing tool that can add significant contextual information to our understanding of the effectiveness of ECE programs. Implementation research might also help reveal how issues such as race, gender, class, and linguistic diversity interact with ECE program delivery and, ultimately, with outcomes for young children. These historically intractable issues may be central to understanding the relationship between populations most in need of services and specific program components or constellations of program components.



**Our ability to achieve better results for young children rests on a more nuanced understanding of how programs are being implemented and the differential impacts on subgroups of children.**

For over 100 years and under changing organizational structures, the Foundation for Child Development has supported research on the well-being of young children. We hope this volume carries on the Foundation's tradition of working to fill gaps in research and making research more relevant and useful to policymakers and practitioners. Our ultimate goal is always to ensure that every child reaches their full potential.

## ACKNOWLEDGEMENTS

---

The Foundation for Child Development would like to extend a special thank you to the authors of this volume for their time, dedication, and contribution to this work. Thanks to everyone at Neimand Collaborative, Armando Molina Bou, Ishaba Haque, Sarah Hutchinson, and Rich Neimand, who made the production and design of this volume possible. Thank you, Jon Wallace and MJ Devaney for your editing expertise.

## References

---

- Brooks-Gunn, J. (2003). Do you believe in magic? What we can expect from early childhood intervention programs. *Social Policy Report*, 27(1), 3-15.
- Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology*, 37(2), 231-242.
- Heckman, J., Moon, S. H., Pinto, R., Savelyev, P., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, 94(1-2), 114-128.
- Lipsey, M., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee prekindergarten program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly*, 45(4), 155-176.
- Lipsey, M., Farran, D. C., & Hofer, K. G. (2015). *A randomized control trial of the effects of a statewide voluntary prekindergarten program on children's skills and behaviors through third grade* (Research report). <https://files.eric.ed.gov/fulltext/ED566664.pdf>
- Ramey, C. T., Campbell, F. A., Burchinal, M., Skinner, M. L., Gardner, D. M., & Ramey, S. L. (2000). Persistent effects of early intervention on high-risk children and their mothers. *Applied Developmental Science*, 4(1), 2-14.
- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The High/Scope Perry Preschool study through age 40*. [http://nieer.org/wp-content/uploads/2014/09/specialsummary\\_rev2011\\_02\\_2.pdf](http://nieer.org/wp-content/uploads/2014/09/specialsummary_rev2011_02_2.pdf)

SECTION 1

WHAT DOES RESEARCH TELL  
US ABOUT EFFECTIVENESS  
AND IMPLEMENTATION OF ECE  
PROGRAMS ACROSS THE  
BIRTH-TO-EIGHT CONTINUUM?

## **IN SECTION 1:**

---

### **Chapter 1: What Does Research Tell Us About ECE Programs?**

By Margaret R. Burchinal, Ph.D., University of North Carolina at Chapel Hill and Dale C. Farran, Ph.D., Vanderbilt University

### **Chapter 2: What Are Reasonable Expectations for ECE Program Effectiveness?**

By Jeanne Brooks-Gunn, Ph.D., Teachers College and College of Physicians and Surgeons, Columbia University and Sarah Lazzeroni, Teachers College, Columbia University

### **Chapter 3: Using a Social Determinants of Early Learning Framework to Eliminate Educational Disparities and Opportunity Gaps.**

By Iheoma U. Iruka, Ph.D., HighScope Educational Research Foundation

**SECTION 1, CHAPTER 1**

**WHAT DOES RESEARCH TELL US  
ABOUT ECE PROGRAMS?**

---

Margaret R. Burchinal, Ph.D., University of North Carolina at Chapel Hill

Dale C. Farran, Ph.D., Vanderbilt University



## INTRODUCTION

---

Early care and education (ECE) now plays an integral role in early development, so it is important to understand how ECE affects children's learning and development. This chapter describes the extensive literature relating ECE quality and programs to both short- and long-term development. The findings from these ECE research and evaluation studies are contrasted and discussed in the context of factors that limit current ECE programs and policies from achieving the goal of promoting positive short- and long-term outcomes for all children.

ECE serves two primary functions: supporting parental employment and promoting positive cognitive and social development to reduce achievement gaps during the school years (Burchinal, Magnuson, Powell, & Hong, 2015). Its first function is to care for very young children while their parents work; in the United States (U.S.) and much of the world, most men and over two-thirds of women are employed outside the home (OECD, 2018). At this time, over 80% of preschoolers (three- to five-year-olds) and 35% of infants (zero- to two-year-olds) attend ECE programs (OECD, 2018). Many other children, especially infants, experience out-of-home care by relatives (Burchinal et al., 2015). Parents' decisions about ECE, as well as the options open to them, depend on cultural norms (Lamb, 1998). In northern Europe, for example, ECE is viewed as a community responsibility. Parents are offered generous, government-subsidized family leave and low-cost, high-quality child care (Waldfogel, Han, & Brooks-Gunn, 2002). In contrast, in the U.S. and much of the rest of the world, childrearing is viewed as primarily the family's responsibility. Most parents choose from a range of options in the private market, especially for infants, toddlers, and young preschoolers (Waldfogel et al., 2002). As a result, ECE has mostly remained a family responsibility in the U.S. (National Academies of Sciences, Engineering, and Medicine, 2018).

ECE's second function is to promote children's cognitive and social development before they enter elementary school (Burchinal et al., 2015). Experimental early intervention studies conducted prior to 1980 demonstrated that ECE could have long-term impacts on low-income children's educational and labor-market success (Heckman, 2011). Accordingly, ECE became a primary policy mechanism for addressing concerns that some children, particularly low-income children, arrive at school unprepared to succeed in elementary school, and that differences in school readiness have lasting consequences (Burchinal et al., 2015). It is argued that ECE programs generate benefits not only to participants but also to the economic and social health of communities (Barnett & Masse, 2007; Heckman & Masterov, 2007; Magnuson, Ruhm, & Waldfogel, 2007; Putnam, Frederick, & Snellman, 2012). As a consequence, a variety of programs have been publicly funded to increase access to high-quality ECE, including the federally funded Head Start program primarily for low-income children, state-funded pre-k programs typically for low-income children, and state Quality Rating and Improvement Systems (QRIS) designed to improve access to high-quality ECE for all children (Barnett, 2013).

The large ECE research literature has answered important questions about the quality of ECE programs and their impact on young children’s development. But ECE research has not fully examined implementation of programs or policies to determine how components, contexts, fidelity, and target populations relate to child outcomes. It is clear that young children thrive when caregivers are responsive and sensitive in their interactions and stimulate learning by providing and scaffolding age-appropriate activities. But only some research has asked whether specific program or ECE quality indicators relate to child outcomes differently for children of different races/ethnicities, social classes, or home languages or even whether different aspects of the ECE experience promote different child outcomes. Furthermore, most ECE research is based on a theoretical model that posits that structural quality (e.g., characteristics such as teacher education and ratio of children to adults) lays the foundation for process quality (i.e., the frequency and quality of interactions between caregivers and children), and that it is process quality that impacts child outcomes. But the evidence supporting this model using current measures of structural and process quality is quite limited. Thus, we do not know enough about what works (or not), for whom, and under what conditions in promoting which skills for young children. This volume addresses all of these questions and this chapter discusses the research regarding these issues.

## WHAT’S WORKING IN ECE PROGRAMS?

---

ECE’s short-term impacts on early learning and development have been measured in several ways. One set of studies has examined associations between indicators of ECE quality, defined in various ways, and child outcomes; another set of studies has evaluated specific types of ECE, including early intervention programs and publicly funded programs and initiatives; and still other studies have focused on specific instructional practices and curricula. The magnitude of ECE’s estimated immediate impacts varies widely both within and between these sets of studies.

### ► ECE quality and child outcomes

Developmental theories suggest that ECE influences children’s learning and development through the quality of relationships between caregivers and children and opportunities to learn through hands-on, age-appropriate activities that adults scaffold (see Burchinal et al., 2015 & Hamre, 2014 for details). Attachment theory postulates that frequent, warm, and sensitive interactions with caregivers allow children to engage meaningfully with objects and people in their environment (Ainsworth, Blehar, Waters, & Wall, 1978; Howes & Spieker, 2008). Piaget’s constructivist developmental theory argues that early cognitive development requires children to actively engage with objects and people to learn (Gopnik, Meltzoff, & Kuhl, 1999). Vygotsky’s social-cultural theory describes how caregiver scaffolding aids learning (Vygotsky, 2001). Bronfenbrenner’s ecological theory emphasizes the critical role of primary caregivers at home and in ECE, as well as the continuity and connections between the two contexts (Bronfenbrenner & Morris, 2006).

Definitions of ECE have evolved from these theories and from developmental research. “Process quality” describes the quality of factors that directly affect children in ECE, either through the frequency and quality of their interactions with caregivers or through their access to engaging and informative activities. Certain program and teacher characteristics are thought to promote process quality, including factors such as caregiver education and training, child/adult ratios and group size, and curriculum. These “structural quality” factors indirectly affect children through their presumed impact on process quality. Simplistically represented, the following model suggests these causal links (NICHD ECCRN, 2002):



**Structural quality.** Structural quality is thought to be important because it provides caregivers with the skills, knowledge, and opportunity to provide the high process quality that can improve child outcomes (NICHD ECCRN, 1999, 2002). Structural quality indicators include the caregivers’ education and training, wages and benefits, the ratio of children to caregivers, the number of children in a setting, program leadership and administration, and parental involvement (Build Initiative & Child Trends, 2014; Burchinal, Tarullo, & Zaslow, 2016).

Research indicates that process quality is higher when structural quality is higher. Earlier research found that teacher education, teacher training, ratio of children to adults, group size, caregiver wages, and administrator experience and communication style had moderate-to-strong associations with both global environmental quality (Bloom & Sheerer, 1992; Burchinal et al., 2000b; Phillipsen, Burchinal, Howes, & Cryer, 1997) and ratings of teacher-child relationship sensitivity (NICHD ECCRN 1999, 2002a). But these associations have not always been observed (Mashburn et al., 2008; Pianta et al., 2005).

Whereas the pathway from structural quality indicators through process quality to child outcomes has been supported in at least one study (NICHD ECCRN, 2002a), many studies have examined associations between structural quality and child outcomes. They looked at the direct pathways from structural quality to child outcomes, in part because compared to process quality, the structural quality indicators can be more easily monitored and therefore are easier to use in licensing or performance monitoring of ECE programs. These studies provide some limited evidence of associations. In early

studies, preschool children's outcomes were modestly better when their teachers had more education (Burchinal et al., 2000b; Phillipsen et al., 1997; NICHD ECCRN, 2002a) and classrooms had fewer children per teacher (NICHD ECCRN, 2002a; Phillipsen et al., 1997). When the number of children in a preschool classroom was larger, behavior problems were reported more frequently (McCartney et al., 2010). In addition to individual studies, meta-analyses using large child care studies suggest that children's skills levels are higher when caregivers receive training, especially with intensive training or training aligned with a rigorous curriculum (Fukkink & Lont, 2007), and when teachers and administrators have more education (Burchinal et al., 2016). But they are not higher when child-adult ratios or group size are smaller (Burchinal et al., 2016). And even when the associations between structural quality indicators and child outcomes in the recent studies were statistically significant, their magnitude was quite modest—most effect sizes were .10 or smaller. In summary, some but not all evidence suggests that some structural quality indicators are very modestly related to some child outcomes.

**Process quality.** All ECE models assume that the quality of interactions between caregivers and children (process quality) determines the extent to which ECE experiences are positive for children and are the processes through which ECE impacts early learning and development (Burchinal et al., 2015). Process quality reflects the extent to which caregivers are responsive and sensitive with the children in their care, provide stimulating activities, and scaffold early learning and development (Hamre, 2014). There are two widely used tools for measuring ECE quality. The Environmental Rating Scales (ERS) (Harms, Clifford, & Cryer, 2005) focuses on the extent to which children have hands-on opportunities for learning and on the level of caregiver scaffolding during those activities. The Classroom Assessment Scoring System (CLASS) (Pianta, La Paro, & Hamre, 2008) describes the quality of the teacher-child relationship. The ERS focuses on children's access to a variety of age-appropriate activities and if/how caregivers engage with them during those activities. It includes the Early Childhood Environment Rating Scale (ECERS) (Harms, Clifford, & Cryer, 2005) to describe the quality of preschool center care, the Infant-Toddler Environmental Rating Scale (ITERS) (Harms, Cryer, & Clifford, 2003) to describe the quality of infant/toddler center care, and the Family Day Care Environment Rating Scale (FDCERS) (Harms, Cryer, & Clifford, 2007) to describe home-based care. These measures emphasize the types and variety of activities provided, the extent to which the child is an active participant in the learning process, and the extent to which adults engage with children in those activities. Each one also assesses the provider's sensitivity and responsiveness, health-related practices and the safety of the setting, and classroom-management practices. According to these measures, a high-quality classroom has at least five different interest centers, conversations during meal and snack time, a wide selection of books that are read in formal class activities and in informal interactions with the teacher, and activities that encourage children to think, talk, and reason about their experiences (Harms et al., 2005).

The CLASS focuses on the quality of interactions between children and their caregivers and the level of positive classroom management. It is an extension of a scale, the Observational Rating of the Childcare Environment (ORCE), developed by the NICHD Study of Early Child Care and Youth Development (NICHD ECCRN, 1997). It rates caregivers' warmth and sensitivity and the instructional support they provide, as well as the degree to which their classroom management is positive and effective. According to this measure, teachers in high-quality classrooms have frequent, warm, and responsive interactions with children. The teacher attends to each child, individualizing feedback to match his or her skill level. The teachers talk frequently with each student in multi-turn conversations in which the adult elaborates on the students' responses by asking open-ended questions (Hamre, 2014).

The associations between these process-quality measures and child outcomes have been examined extensively. The earliest studies reported moderate associations, typically between the ECERS and child outcomes (e.g., Burchinal et al., 2000a; Clarke-Stewart, 1998; Howes, Rodning, Galluzzo, & Myers, 1988; Peisner-Feinberg & Burchinal, 1997; McCartney, 1984). These studies were criticized, however, because they included only a few demographic characteristics and therefore failed to account for potential differences in the families that selected different quality levels of ECE for their children. That is, more advantaged parents choose higher-quality care and have children with higher levels of developmental skills, so the children's higher skill levels may have more to do with family advantage than with ECE quality (Duncan & Magnuson, 2004). The next set of studies included extensive family and child covariates and yielded statistically significant but modest associations between child outcomes and ECE quality (e.g., Howes et al., 2008; Mashburn et al., 2008; NICHD ECCRN, 2002; Votruba-Drzal, Coley, & Chase-Lansdale, 2004). Some of these studies asked whether a certain level of quality (i.e., a threshold) was necessary for quality to improve child outcomes. Some evidence of quality thresholds for the CLASS domain scores emerged, but it was inconsistent. And the associations between quality and outcomes remained modest even above the threshold (Burchinal et al., 2010; 2014; 2016; Hatfield et al., 2015; Weiland, Ulvestad, Sachs, & Yoshikawa, 2013). Most recently, several meta-analyses that reanalyzed large ECE studies also found reliable but very small associations with some child outcomes, with effect sizes of around .05 (Keys et al., 2013; Burchinal et al., 2016).

Specific aspects of ECE quality appear to enhance children's early development. Preschoolers showed modest but significant gains in academic and social skills when they experienced more frequent, warm, and responsive interactions with caregivers (Mashburn et al., 2008; NICHD ECCRN, 2002; Raver et al., 2011). Gains in academic skills are modestly larger when instruction includes detailed feedback, and sequenced and elaborated support for learning (Howes et al., 2008; Mashburn et al., 2008). Language and academic skills were higher when caregivers encouraged children to talk and engaged in multi-turn conversations that elaborated on a given topic (Justice, Mashburn, Pence, & Wiggins, 2008; Wasik &

Hindman, 2011). Finally, gains in language and social skills were larger when children were offered a wide range of age-appropriate activities such as reading with adults, pretend play with peers, and play with books, blocks, water, and sand, demonstrated gains in language and social skills (Sylva et al., 2012).

**Policy applications of the ECE model.** The major policy initiative designed to improve access to high quality care was developed using the conceptual model relating structural quality to process quality to child outcomes was used to develop QRIS programs. States and localities developed these ratings systems using structural- and process-quality indicators to describe the quality of participating ECE programs, and provided incentives and professional development to enrolled programs. All QRIS ratings include measures of process quality and structural quality (e.g., caregiver education and training, and group size or child-adult ratio), and many include measures of parental involvement, inclusion of children with special needs, and practices that align programs with the family practices for children who come from diverse backgrounds or who speak a language other than English at home (Build Initiative & Child Trends, 2014). Validation studies of QRIS systems in many states have found that ECE programs at higher QRIS-quality tiers had higher process quality as indicated by higher ERS or CLASS scores, providing reassuring validation for the rating systems (e.g., Lipscomb, Weber, Green, & Patterson, 2016; Tout, Cleveland, Li, Starr, Soli, & Bultnick, 2016; Yazejian et al., 2017). But these validation studies yielded little to no evidence of higher skills levels among children who attend programs at higher quality tiers, raising questions about the pathways from process quality to child outcomes in the ECE model underlying the QRIS systems (Karoly, Schwartz, Setodji, & Haas, 2016; Sabol & Pianta, 2015; Soliday Hong et al., 2015; Thornburg, Mayfield, Hawks, & Fuger, 2009; Yazejian et al., 2017; Zellman, Perlman, Le, & Setodji, 2008).

### ► **Child outcomes and ECE instructional practices and programs**

Other studies have examined the short-term impacts of specific early childhood teaching practices and ECE programs. A meta-analysis of all randomized clinical trials of early childhood interventions yielded an average effect size of about .35 for most of these ECE programs and practices (Duncan & Magnuson, 2013). Stronger impacts were found for studies of intensive curricula with scope and sequence. Evidence-based curricula, when combined with aligned training or coaching, were related to larger gains in children's literacy skills.

**Teaching practices.** Numerous ECE curricula have been developed and evaluated. Collectively, they demonstrate that a focus on teaching practices and aligned professional development can have substantial impacts on child development across a number of developmental domains. Examples include: a language curriculum with an effect size of .27 (Wasik & Hindman, 2011); a literacy professional development program with effect sizes of .91 to .99 (Powell, Diamond, Burchinal, & Koehler, 2010); a math curriculum with effects sizes of .47 to 1.07 (Clements & Sarama, 2008); and a social-emotional

**Numerous ECE curricula have been developed and evaluated. Collectively, they demonstrate that a focus on teaching practices and aligned professional development can have substantial impacts on child development across a number of developmental domains.**

learning curriculum with an effect size of .63 (Raver et al., 2008). Integrating several evidence-based curricula has also had modest-to-large impacts on child outcomes. For example, the Boston Public Schools Universal Pre-K program integrated evidence-based literacy and math curricula and children showed moderate-to-large gains in those content areas (effect sizes of .45 to .82, respectively), as well as more modest gains in executive functioning (EF) (effect sizes of .21 to .28; Weiland & Yoshikawa, 2013).

*ECE programs.* Between 1960 and 1980, ECE intervention programs demonstrated large short-term outcomes. These include the Perry Preschool/HighScope program (Cunha & Heckman, 2007) and Chicago Parent-Child Centers (Reynolds, Magnuson, & Ou, 2010), which combined child care and parenting programs for preschoolers and their mothers, and the Abecedarian Project (Campbell et al., 2012), which provided full-time child care and onsite medical care from infancy to kindergarten. Abecedarian yielded large short-term impacts on cognitive development, and the other projects obtained moderate short-term impacts on cognitive and social outcomes.

Statistically rigorous evaluations of publicly funded programs have also found modest-to-large short-term impacts for Head Start and some state pre-K programs. An experimental study of Head Start, the federally funded program for low-income children, yielded modest impacts at the end of one year of the program (Puma, Bell, Cook, & Heid, 2010). State pre-K programs vary widely from state to state (Barnett, 2013), so it is not surprising that estimates of immediate impact vary from nil to very large (Phillips et al., 2017). Children attending the pre-K programs that meet most professional guidelines tend to show moderate-to-large immediate gains, with the largest gains among dual language learners and children from low-income families (Phillips et al., 2017). Most pre-K evaluations report statistically significant moderate-to-large impacts on rote reading and math skills, but smaller or no reliable impacts on language, social skills, and EF (Burchinal, 2017).

Interpreting ECE program evaluations can be complicated by the timing of program implementation and methodological issues. The studies with the largest short-term impacts are the small, experimental ones conducted in the 1960s and early '70s. Duncan and Magnuson (2013) warn that generalizing the results from those findings to today's programs is problematic because the comparison groups in these studies are very different from the comparison groups of today. In the earlier studies, the comparison groups consisted primarily of low-income children who stayed home before kindergarten; the comparison

group in studies conducted in the past year consists of low-income children who attend other types of center care. Given that center care appears to be beneficial, especially for low-income children (Magnuson et al., 2007), this change in the counterfactual makes it more difficult to detect impacts (Duncan & Magnuson, 2013). For example, immediate Head Start impacts appear to be much larger if Head Start children are compared to children who did not attend center care, especially if they spoke Spanish at home (Bloom & Weiland, 2015). In addition, quasi-experimental studies used to evaluate pre-K programs rely heavily on statistical assumptions in estimating pre-K impacts, and those impacts may be inflated due to violations in those assumptions, such as differential attrition in the treated group (Phillips et al., 2017). The evaluation of Boston's pre-K program, which attempted to address differential attrition, yielded somewhat smaller effect sizes than those reported in some of the other evaluations (Minervino, 2014).

► **Potential reasons for larger ECE effects in studies of programs and practices than in studies of quality**

In summary, the studies relating process and structural ECE quality to short-term child outcomes report very small associations, whereas at least some of the studies of programs and curricula report moderate-to-large associations. These findings challenge our current models of how ECE influences child outcomes, which argue that process quality—the quality of teacher-child interactions and access to hands-on learning experiences—determines children's learning and development in ECE, and that other ECE factors, such as instructional practices and programs, have their impacts through improving process quality (e.g., Hamre, 2014). Despite the widespread belief that when ECE programs positively impact child outcomes these impacts occur because the programs are of high quality, little evidence actually links program efficacy to measures of process quality. Furthermore, the impacts of the effective programs are much larger than observed associations between process quality and child outcomes (Burchinal, 2017). This raises questions about whether current quality measures are adequate or whether our ECE models need to be expanded (Burchinal, 2017).

**Psychometric issues.** Limited variability on existing scales has created psychometric problems. Designed to be aspirational, the widely used ERS and CLASS systems measure a full range of very bad (i.e., a rating of 1) to very good (i.e., a rating of 7) quality on each item. Consequently, most classrooms tend to be rated somewhere in the middle, within a small range of the overall scale. The small standard deviation for each item tends to be less than 1 point. Raters are certified as reliable when 80 to 85% of their item scores are within 1 point of the trainer's rating or the gold standard. This creates large variability among raters, and inter-rater variability often accounts for more than 25% of total variance in classroom-quality ratings (Burchinal, 2017).

**Restricted scope of ECE quality measures.** Larger impacts in evaluations of curricula and pre-K programs with a stronger instructional component suggest that the quality of intentional teaching needs to be measured more carefully (Burchinal, 2017; Yoshikawa et al., 2013). Professional development randomized clinical trials that improved the quality of teacher-child interactions as measured by the CLASS failed to improve child outcomes (Pianta et al., 2017; Yoshikawa et al., 2015), suggesting that improving process quality as measured by the CLASS may not be sufficient to change academic skills in particular. Because specific curricula and pre-K programs show much larger impacts, ECE quality measures may need to focus more on the frequency and quality of intentional teaching. Furthermore, it may be necessary to examine instruction within content areas because teachers may differ in their ability to cover subjects like literacy, math, and science.

Recently, several measures have shown promise for expanding the measurement of ECE quality. They involve behavioral counts rather than ratings, and they vary in terms of whether the unit of observation is the teacher or multiple children in the classroom. Connor et al., 2011 developed an integrated system involving child monitoring, classroom observations, and instruction that has been shown to substantially improve reading skills in early elementary school; a preschool version is in the works.

Observational measures that describe how children spend their time and how teachers interact with them appear promising. One, the Snapshot (Ritchie, Weiser, Kraft-Sayre, & Howes, 2001), describes how much time individual children spend in different types of activities in terms of content area and instructional format. When districts used the Snapshot to create pre-k to third-grade programs, child outcomes improved and parents became more involved (Manship, Farber, Smith, & Drummond, 2016). Two other measures, the Language Interaction Snapshot (LISn) (Sprachman, Caspe, & Atkins-Burnett, 2009) and Observation Measures of Language and Literacy Instruction (OMLIT) (Goodson et al., 2004), describe the frequency and quality of linguistic interactions in ECE classrooms. Children who have more frequent and complex linguistic interactions with their teachers showed moderate to large gains in their language skills (Abt Associates, 2007). The Child Observation in Preschool/Teacher Observation in Preschool (COP/TOP) (Farran & Son-Yarborough, 2001; Bilbrey, Vorhaus, Farran, & Shufelt, 2010) measures how much and to whom the teacher talks and listens, the types of tasks in which the teacher or assistant is engaged, the level of ongoing instruction or assessment, the content area, and the tone of the interactions. Results from this measure have been associated with both short- and long-term gains in self-regulation (Fuhs, Farran & Turner, 2013; Spivak & Farran, 2016) as well as academic outcomes (Farran et al., 2017).

## LONG-TERM ECE IMPACTS ON CHILD OUTCOMES

---

Research on the long-term impact of ECE quality, instruction, and program has yielded mixed findings. Early studies demonstrated important long-term impacts into adulthood on education, employment, family formation, and health (Campbell et al., 2012; 2014; Cunha & Heckman, 2007; Reynolds et al., 2010). On the other hand, later studies of process quality, instruction, and programs have suggested that impacts may fade over time.

Three large studies of process quality documented very small but reliable associations between preschool quality and child outcomes in elementary school (Belsky et al., 2007; Peisner-Feinberg et al., 2001; Sylva et al., 2012) and high school (Vandell et al., 2010). Follow-up studies of the pre-k programs indicate smaller but still significant, longer-term effects for some of the most rigorous programs (Phillips et al., 2017). Long-term quasi-experimental studies suggest that Head Start has positive impacts into adulthood (Ludwig & Miller, 2007). Yet many studies do not show long-term gains. The meta-analysis of all early childhood interventions reported that the average impact declined during the elementary years and was not significantly different from zero by the end of elementary school (Duncan & Magnuson, 2013). The follow-up study of the experimental evaluation of Head Start indicated that all impacts disappeared early in elementary school (Puma et al., 2012). One of the most rigorous evaluations of any pre-K program, that of the Tennessee Pre-K Program, showed negative impacts on outcomes in third grade (Farran & Lipsey, 2015).

*Inadequate attention to some school-readiness skills.* The child outcomes that ECE seeks to improve have changed over time. Early programs, such as Abecedarian and HighScope, focused on improving general knowledge and language skills. Teachers engaged in frequent conversations with children and, through conversations and activities, actively scaffolded children's learning. (Ramey & Ramey, 1998; Lazar et al., 1982). Head Start originally focused on improving nutrition and social skills to provide the basis for success in school (OHS, 2018). Head Start and most child care programs added a primary on academic skills starting about 20 years ago based on evidence that having these academic skills at entry to school was the basis for school-age academic achievement (Burchinal et al., 2015). Thus, it is not surprising that, as described above, the immediate impacts of ECE programs tend to be on academic skills, rather than language, EF, or social skills.

The focus on teaching basic reading and math skills in preschool programs likely contributes to fade-out because it appears these skills are also taught in kindergarten. Despite the fact that more than three-fourths of children in a nationally representative study entered kindergarten with basic literacy and math skills, kindergarten teachers spend most of their time teaching those skills (Claessens, Engel, & Curran, 2013; Engel, Claessens, & Finch, 2013). Indeed, the only children who made substantial gains in literacy in during kindergarten had not mastered those skills prior to entry to school. Thus, it is likely that the lack

of continuity between instruction in preschool and kindergarten that accounts for much of the fade-out in academic skills.

In addition, focusing on academic skills may contribute to fade-out if other skills are important academically and socially during the school years. A comprehensive review (National Research Council, 2008) differentiated between rote skills, such as basic literacy and numeracy learned through direct instruction, and higher-order skills, such as oral language and EF acquired through extended, scaffolded interactions with caregivers. Evidence suggests that higher-order skills at school entry predict acquisition of later academic skills better than basic skills (Blair & Raver, 2012; Snow & Oh, 2010). Other studies have also related multiple school-readiness skills to academic and social skills in elementary school. The school-readiness skills most consistently related to school-age skills were

**Along with the early intervention studies, these studies suggest that doing more to promote general knowledge, language, EF, and self-regulation might give children skills that improve their academic and social outcomes during the school years.**

language (Pace et al., 2017), general knowledge (Grissmer et al., 2010), and self-regulation and EF skills (Fuhs, Nesbitt, Farran, & Dong, 2014).

Though math skills have also been found to predict subsequent reading and math outcomes (Duncan et al., 2007), later work suggests that including cognitive skills in the analyses would have yielded different conclusions (Bailey, Watts, Littlefield, & Geary, 2014; Grissmer et al., 2010). Along with the early intervention studies, these studies suggest that

doing more to promote general knowledge, language, EF, and self-regulation might give children skills that improve their academic and social outcomes during the school years.

An important question to answer is the degree to which fade-out is related to a lack of alignment in reading and math instruction from preschool to kindergarten, or to the focus on academic rather than higher-order skills in preschool. To the extent that kindergarten teachers teach skills that children learned during their preschool years, it is not possible now to determine the relative contribution of these two explanations.

**Characteristics of preschool programs.** Preschool programs operate in ways that may make it difficult to meet expectations regarding child outcomes. These programs typically follow the school model of offering up to six hours of care per day for up to nine months per year. The opportunities for learning during those six hours are limited by the time required for naps, toileting, and meals, and in the worst programs children spend much of their time transitioning among activities (Early et al., 2006). Many preschool programs focus on large-group, didactic instruction that is not developmentally appropriate for preschoolers (Farran & Lipsey, 2015).

Preservice preparation for ECE teachers, including college and certification programs, is a matter of deep concern; problems include a lack of focus on producing ECE teachers and a lack of consistency and rigor in courses, teaching staff, and certification requirements (Early et al., 2007). Similarly, we lack evidence that in-service training programs are effective, despite huge expenditures on professional development and technical assistance.

Last, preschool teachers' low salaries in both community settings and several state-funded programs limit ECE quality by determining who becomes and remains a preschool teacher. Wages are low because parents typically pay for community-based ECE, and most parents cannot afford to pay the higher fees that allow for higher wages for teachers. Child care vouchers for low-income children to attend ECE while their mothers work or go to school are often indexed to average fees in the community (Burchinal et al., 2015). Public programs, such as Head Start and state pre-k, often offer slightly higher salaries, but pay is still typically below that of certified elementary education teachers (Burchinal et al., 2015). Consequently, it is difficult to recruit and retain highly qualified ECE teachers, which constrains ECE quality in community-based organizations and publicly funded programs (National Academies of Sciences, Engineering, and Medicine, 2018).

## WHAT NEEDS TO BE UNDERSTOOD ABOUT ECE?

---

We need to understand many other issues if we are to meet ECE's promise to ensure that children enter school ready to succeed in primary school and beyond. One such issue is the extent to which children's race/ethnicity and home language may require attention to different or additional factors (McCabe et al., 2013). For example, there is considerable evidence that dual language learners benefit from practices that promote their first language while they learn their second language, especially during early childhood (Espinosa, 2013; McCabe et al., 2013). Evidence is mixed regarding the degree to which having an ECE provider from the same ethnicity/race improves young children's ECE experiences, but developmental theories suggest that continuity between home and ECE should make it easier for children to develop and learn (Gillanders, Iruka, Ritchie, & Cobb, 2012; Schick, 2014).

We also need to pay more attention to practices that facilitate the transition to elementary school (transition papers) and continuity of care from preschool through third grade (Bogard & Takanishi, 2005; Reynolds et al., 2010; Stipek et al., 2017). Transition activities like communication between the preschool and kindergarten teachers improve child outcomes during kindergarten (Ahtola et al., 2010; LoCasale-Crouch, Mashburn, Downer, & Pianta, 2008). Continuity in expectations and learning opportunities between pre-K and the first four years of elementary school helps children both maintain preschool gains and make larger gains in elementary school (Reynolds et al., 2010). Careful alignment among evidence-based instruction, assessment, and professional development within and between years appears to maintain gains in elementary school (Bryk et al., 2010).

Another area that needs more attention is identifying which school-readiness skills promote long-term development and which ECE practices promote those skills. Current ECE quality models assume that children acquire cognitive, academic, and social skills when they experience high levels of process quality, but the models do not specify how quality experiences promote specific skills. The fact that we see much larger impacts on outcomes in studies of specific curricula (Duncan & Magnuson, 2013) than in studies of ECE quality (Burchinal, 2017) suggests that ECE can produce substantial gains in specific outcomes when it promotes those outcomes with evidence-based practices. Once evidence identifies which school-readiness skills are related to which school-age academic and social outcomes, we then need to identify ECE instructional practices that promote those skills. It is also important to ask whether those practices vary for children from different ethnicities, social classes, and home languages, and to adapt instructional practices accordingly. We suspect that evidence-based intentional instruction and aligned professional development will focus on teacher-scaffolded learning through rich, multi-turn conversations and sequenced, hands-on activities designed to promote general knowledge, language, EF and self-regulation among young children.

Last but not least, current policies rely primarily on center-based preschool programs that begin at ages three-to-four (e.g., Head Start or state pre-K) to address income and racial achievement gaps, despite clear evidence that a child's first three years are critical for building these foundational skills. By two-to-three years of age, we already see large gaps in language and cognitive skills between children from low- and higher-income families and between children of color and white children (Halle et al., 2009). Preschool programs like Head Start or pre-K can close, but not eliminate, those gaps (Burchinal et al., 2015; Phillips et al., 2017; Yoshikawa et al., 2013). Children's experiences as infants and toddlers at home and in ECE influence their cognitive, academic, and social skills at entry to preschool, so ensuring that children have access high-quality child care during infancy, as well as during the preschool years, may help narrow these gaps (Li et al., 2013).

## CONCLUSION

---

**The field focuses on current measures of ECE quality despite their very modest associations with child outcomes, rather than on the evidence-based curricula or specific types of ECE programs that have much larger impacts on child outcomes.**

Early care and education can improve young children’s academic and social skills, with some evidence of long-term impacts during the school years and into adulthood. Yet there are many reasons to believe most ECE programs could be much more effective. The field focuses on current measures of ECE quality despite their very modest associations with child outcomes, rather than on the evidence-based curricula or specific types of ECE

programs that have much larger impacts on child outcomes. Identifying which preschool skills promote the acquisition of which specific school-age skills should lead to greater focus on promoting those skills in ECE. Models that pay greater attention to which specific instructional practices improve those skills are likely to be more successful than our current models when it comes to achieving ECE’s promise of promoting long-term development for all children.

## References

---

- Abt Associates, Inc. (2007, March). *Findings from Project Upgrade in Miami-Dade County*. Cambridge, MA.
- Ahtola, A., Silinskas, G., Poikonen, P., Kontoniemi, M., Niemi, P., & Nurmi, J. (2011). Transition to formal schooling: Do transition practices matter for academic performance? *Early Childhood Research Quarterly*, *26*, 295-302. doi: 10.1016/j.ecresq.2010.12.002
- Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. N. (1978). *Patterns of attachment: A psychological study of the strange situation*. Hillsdale, NJ: Erlbaum.
- Bailey, D. H., Watts, T. W., Littlefield, A. K., & Geary, D. C. (2014). State and trait effects on individual differences in children's mathematical development. *Psychological Science*, *25*(11), 2017-2026. doi: 10.1177/0956797614547539
- Barnett, W. S. (2013). Expanding access to quality pre-K is sound public policy. *National Institute for Early Education Research*, *18*.
- Barnett, W. S., & Masse, L. N. (2007). Comparative benefit-cost analysis of the Abecedarian program and policy implications. *Economics of Education Review*, *26*(1), 113-125. doi: 10.1016/j.econedurev.2005.10.007
- Belsky, J., Vandell, D.L., Burchinal, M., Clarke-Stewart, K.A., McCartney, K., Owen, M.T., and the NICHD ECCRN. (2007). Are there long-term effects of early child care? *Child Development*, *78*(2), 681-701.
- Billbrej, C., Vorhaus, E., Farran, D. C., & Shufelt, S. (2010). *Teacher observation in preschool: Tools of the Mind adaptation*. Nashville, TN: Peabody Research Institute.
- Blair, C. & Raver, C. C. (2012). Child development in the context of adversity: Experiential canalization of brain and behavior. *American Psychologist*, *67*(4), 309-318. doi: 10.1037/a0027493
- Bloom, P. J., & Sheerer, M. (1992). The effect of leadership training on child care program quality. *Early Childhood Research Quarterly*, *7*(4), 579-594.
- 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.
- Bogard, K., & Takanishi, R. (2005). PK-3: An aligned and coordinated approach to education for children 3 to 8 years old. *Society for Research in Child Development Social Policy Reports*, *19*(3), 1-23.
- Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner (Ed.), *Theoretical models of human development*. In *Handbook of Child Psychology*. (Vol. 1, 6th ed., pp. 793-828). Editors-in-Chief: W. Damon & R. M. Lerner. Hoboken, NJ: Wiley.
- Bryk, A. S., Sebring, P. B., Allensworth, E., Easton, J. Q., & Luppescu, S. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago: University of Chicago Press.
- Build Initiative & Child Trends. (2014). A Catalog and Comparison of Quality Rating and Improvement Systems (QRIS) [Data System]. Retrieved from the QRIS Compendium website at <http://qriscompendium.org> on May 14, 2016.
- Burchinal, M. (2017). Measuring early care and education quality. *Child Development Perspectives*, *12*, 3-9. doi: 10.1111/cdep.12260
- Burchinal, M., Magnuson, K., Powell, D., & Hong, S. S. (2015). Early child care and education and child development. In M. Bornstein, R. Lerner, & T. Leventhal (Eds.). In *Handbook of child psychology and developmental science* (Vol. 4, 7th ed., pp. 223-267). Hoboken, NJ: Wiley

## CHAPTER 1 WHAT DOES RESEARCH TELL US ABOUT ECE PROGRAMS?

- Burchinal, M. R., Roberts, J. E., Riggins, R., Zeisel, S. A., Neebe, E., & Bryant, D. (2000a). Relating quality of center-based child care to early cognitive and language development longitudinally. *Child Development, 71*, 339-357. doi: 10.1111/1467-8624.00149
- Burchinal, M. R., Peisner-Feinberg, E., Bryant, D. M., & Clifford, R. (2000b). Children's social and cognitive development and child care quality: Testing for differential associations related to poverty, gender, or ethnicity. *Applied Developmental Science, 4*, 149-165. doi: 10.1207/S1532480XADS0403\_4
- Burchinal, M., Tarullo, L., & Zaslow, M. (2016). *Best practices in creating and adapting Quality Rating and Improvement System (QRIS) rating scales*. OPRE Research Brief #2016-25. Retrieved from [https://www.acf.hhs.gov/sites/default/files/opre/ccepra\\_qris\\_531\\_508compliant\\_66\\_b508.pdf](https://www.acf.hhs.gov/sites/default/files/opre/ccepra_qris_531_508compliant_66_b508.pdf) on March 28, 2018.
- Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly, 25*, 166-176. doi: 10.1016/j.ecresq.2009.10.004
- Burchinal, M., Vernon-Feagans, L., Vitiello, V., & Greenberg, M. (2014). Thresholds in the association between child care quality and child outcomes in rural preschool children. *Early Childhood Research Quarterly, 29*(1), 41-51. doi: 10.1016/j.ecresq.2013.09.004
- Campbell, F. A., Pungello, E. P., Kainz, K., Burchinal, M., Pan, Y., Barbarin, O., Sparling, J. J., & Ramey, C.T. (2012). Adult outcomes as a function of an early childhood educational Program: An Abecedarian Project follow-up. *Developmental Psychology, 48*, 1033-1043. doi: 10.1037/a0026644
- Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. (2014). Early childhood investments substantially boost adult health. *Science, 343*(6178), 1478-1485. doi: 10.1126/science.1248429
- Claessens, A., Engel, M., & Curran, F. C. (2013). Academic content, student learning, and the persistence of preschool effects. *American Education Research Journal, 51*, 403-434. doi: 10.3102/0002831213513634
- Clarke-Stewart, K. A. (1998). Historical shifts and underlying themes in ideas about rearing young children in the United States: Where have we been? Where are we going? *Infant and Child Development, 7*(2), 101-117. doi: 10.1002/(SICI)1099-0917(199806)7:2<101::AID-EDP167>3.0.CO;2-6
- Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal, 45*, 443-494. doi: 10.3102/0002831207312908
- Connor, C. M., Morrison, F. J., Fishman, B. J., Giuliani, S., Luck, M., Underwood, P.,...Schatschneider, C. (2011). Testing the impact of child characteristics X instruction interactions on third graders' reading comprehension by differentiating literacy instruction. *Reading Research Quarterly, 46*, 189-221.
- Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review, 97*(2), 31-47. doi: 10.1257/aer.97.2.31
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P.,...Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428-1446. doi: 10.1037/0012-1649.43.6.1428
- Duncan, G. J., & Magnuson, K. (2004). Individual and parent-based intervention strategies for promoting human capital and positive behavior. Human development across lives and generations: *The potential for change*, 93-135. New York, NY: Cambridge University Press.
- Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. *The Journal of Economic Perspectives, 27*(2), 109-132. doi: 10.1257/jep.27.2.109
- Early, D. M., Bryant, D., Pianta, R., Clifford, R., Burchinal, M., Ritchie, S., Howes, C., & Barbarin, O. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in pre-kindergarten? *Early Childhood Research Quarterly, 21* (2), 174-195. doi: 10.1016/j.ecresq.2006.04.004
- Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., ... & Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78* (2), 558-580. doi: 10.1111/j.1467-8624.2007.01014.x

- Engel, M., Claessens, A., & Finch, M. (2013). Teaching students what they already know? The (mis)alignment between instructional content in mathematics and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis, 35*, 157-178. doi: 10.3102/0162373712461850
- Espinosa, L. M. (2013). *Challenging common myths about teaching young English Language Learners: An update to the seminal 2008 report*. New York, NY: Foundation for Child Development. Retrieved from <https://www.fcd-us.org/prek-3rd-challenging-common-myths-about-dual-language-learners-an-update-to-the-seminal-2008-report/> on March 29, 2018.
- Farran, D. C., & Son-Yarborough, W. (2001). Title I funded preschools as a developmental context for children's play and verbal behaviors. *Early Childhood Research Quarterly, 16*(2), 245-262. psycnet.apa.org/record/2002-06345-007
- Farran, D. C., & Lipsey, M. W. (2015). Expectations of sustained effects from scaled up pre-K: Challenges from the Tennessee study. *Evidence Speaks Reports, 1*(3). Retrieved from <https://www.brookings.edu/research/expectations-of-sustained-effects-from-scaled-up-pre-k-challenges-from-the-tennessee-study/> on March 28, 2018.
- Farran, D.C., Meador, D., Christopher, C., Nesbitt, K. & Billrey, L. (2017). Data-driven improvement in prekindergarten classrooms: Report from a partnership in an urban district. *Child Development, 88*, 1466-1479. doi: 10.1111/cdev.12906
- Fuhs, M., Farran, D., & Turner, K. (2013). Preschool classroom processes as predictors of children's cognitive self-regulation skills development. *School Psychology Quarterly, 28*, 347-359. doi: 10.1037/spq0000031
- Fuhs, M. W., Nesbitt, K. T., Farran, D. C., & Dong, N. (2014). Longitudinal associations between executive functioning and academic skills across content areas. *Developmental Psychology, 50*(6), 1698-1709. <http://doi.org/10.1037/a0036633>
- Fukkink, F., & Lont, A. (2007). Does training matter? A meta-analysis and review of caregiver training studies. *Early Childhood Research Quarterly, 22*, 294-311. doi: 10.1016/j.ecresq.2007.04.005
- Gillanders, C., Iruka, I. U., Ritchie, S., & Cobb, C. (2012). Restructuring and aligning early education opportunities for cultural, language and ethnic minority children. In R. C. Pianta, W. S. Barnett, L. M. Justice, & S. M. Sheridan (Eds.), *Handbook of Early Education* (pp. 111-136). New York, NY: Guilford Publications, Inc.
- Goodson, B. D., C. J. Layzer, and W. C. Smith. (2004). *Observation Measures of Language Literacy Instruction in Early Childhood (OMLIT)*. Cambridge, MA: Abt Associates, Inc.
- Gopnik, A., Meltzoff, A. N., & Kuhl, P. (1999). *The scientist in the crib: What early learning tells us about the mind*. New York, NY: HarperCollins.
- Grissmer, D., Grimm, K. J., Aiyer, S. M., Murrah, W. M., & Steele, J. S. (2010). Fine motor skills and early comprehension of the world: Two new school readiness indicators. *Developmental Psychology, 46*(5), 1008-1017. doi: 10.1037/a0020104
- Halle, T., Forry, N., Hair, E., Perper, K., Wandner, L., Wessel, J., & Vick, J. (2009). *Disparities in early learning and development: Lessons from the Early Childhood Longitudinal Study—Birth Cohort (ECLS-B)* (Child Trends Publication No. 2009-52). Washington, D.C.: Child Trends. Retrieved from <https://www.childtrends.org/wp-content/uploads/2013/05/2009-52DisparitiesELExecSumm.pdf> on March 29, 2019.
- Hamre, B. K. (2014). Teachers' daily interactions with children: An essential ingredient in effective early childhood programs. *Child Development Perspectives, 8*, 223-230. doi: 10.1111/cdep.12090
- Harms, T., Clifford, R. M. & Cryer, D. (2005). *Early Childhood Environment Rating Scale—Revised*. New York, NY: Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R.M. (2003). *Infant/Toddler Environment Rating Scale—Revised*. New York, NY: Teachers College Press.
- Harms, T., Cryer, D. & Clifford, R. M. (2007). *Family Child Care Environment Rating Scale—Revised*. New York, NY: Teachers College Press.

## CHAPTER 1 WHAT DOES RESEARCH TELL US ABOUT ECE PROGRAMS?

- Hatfield, B. E., Burchinal, M. R., Pianta, R. C., & Sideris, J. (2016). Thresholds in the association between quality of teacher–child interactions and preschool children’s school readiness skills. *Early Childhood Research Quarterly, 36*, 561-571. doi: 10.1016/j.ecresq.2015.09.005
- Heckman, J. (2011). The economics of inequality: The value of early childhood education. *American Educator, 31*–47. Retrieved from <http://eric.ed.gov/?id=EJ920516>.
- Heckman, J. J. & Masterov, D. V. (2007). The productivity argument for investing in young children. *Applied Economic Perspectives and Policy, 29*, 446-493. doi: 10.1111/j.1467-9353.2007.00359.x
- Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Ready to learn? Children’s pre-academic achievement in prekindergarten programs. *Early Childhood Research Quarterly, 23*, 27–50. doi: 10.1016/j.ecresq.2007.05.002
- Howes, C., Rodning, C., Galluzzo, D.C., & Myers, L. (1988). Attachment and child care: Relationships with mother and caregiver. *Early Childhood Research Quarterly, 3*, 403-416. doi: 10.1016/0885-2006(88)90037-3
- Howes, C., & Spieker, S. (2008). Attachment relationships in the context of multiple caregivers. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (pp. 317-332). New York, NY: Guilford Press.
- Justice, L. M., Mashburn, A., Pence, K. L., & Wiggins, A. (2008). Experimental evaluation of a preschool language curriculum: Influence on children’s expressive language skills. *Journal of Speech, Language, and Hearing Research, 51*, 983–1001. doi: 10.1044/1092-4388(2008/072)
- Karoly, L. A., Schwartz, H. L., Setodji, C. M., & Haas, A. C. (2016). *Evaluation of Delaware Stars for Early Success*. Santa Monica, CA: Rand Corporation.
- Keys, T. D., Farkas, G., Burchinal, M. R., Duncan, G. J., Vandell, D. L., Li, W., . . . Howes, C. (2013). Preschool center quality and school readiness: Quality effects and variation by demographic and child characteristics. *Child Development, 84*, 1171-1190. doi: 10.1111/cdev.12048
- Lamb, M. (1998). *Nonparental child care: Context, quality, correlates, and consequences*. In I. E. Sigel & K. A. Renninger (Eds.), *Child psychology in practice*. Volume 4 of the *Handbook of child psychology* (pp. 73–133) (5th ed.) Editor-in-Chief: W. Damon. New York, NY: Wiley.
- Lazar, I., Darlington, R., Murray, H., Royce, J., & Snipper, A. (1982). Lasting effects of early education: A report from the Consortium for Longitudinal Studies. *Monographs of the society for research in child development, 47*, (2-3 Serial No. 195).
- Li, W., Farkas, G., Duncan, G. J., Burchinal, M. R., & Vandell, D. L. (2013). Timing of high-quality child care and cognitive, language, and pre-academic development. *Developmental Psychology, 49*, 1440-1451. doi:10.1037/a0030613
- Lipscomb, S. T., Weber, R. B., Green, B. L., & Patterson, L. B. (2016). Quality Rating Improvement System (QRIS).
- LoCasale-Crouch, J., Mashburn, A. J., Downer, J. T., & Pianta, R. C. (2008). Pre-kindergarten teachers’ use of transition practices and children’s adjustment to kindergarten. *Early Childhood Research Quarterly, 23*, 124-139. doi: 10.1016/j.ecresq.2007.06.001
- Ludwig, J., & Miller, D. L. (2007). Does Head Start improve children’s life chances? Evidence from a regression discontinuity design. *The Quarterly Journal of Economics, 122*(1), 159-208. <https://doi.org/10.1162/qjec.122.1.159>
- Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance?. *Economics of Education Review, 26*(1), 33-51. doi: 10.1016/j.econedurev.2005.09.008.
- Manship, K., Farber, J., Smith, C., & Drummond, K. (2016). *Case studies of schools implementing early elementary strategies: Preschool through third grade alignment and differentiated instruction*. Report submitted to the US Department of Education, Office of Planning, Evaluation, and Policy Development, Policy and Program Studies Service. Retrieved from <http://www2.ed.gov/rschstat/eval/implementing-early-strategies/report.pdf>.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D, . . . Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*, 732-749. doi: 10.1111/j.1467-8624.2008.01154.x

McCabe, A., Tamis-LeMonda, C. T., Bronstein, M. H., Brockmeyer Cates, C., ... & Song, L. (2013). Multilingual children: Beyond myths and toward best practices. *Social Policy Report, 27*(4), 1-37.

McCartney, K., Burchinal, M., Clarke-Stewart, A., Bub, K. L., Owen, M. T., & Belsky, J. (2010). Testing a series of causal propositions relating time in child care to children's externalizing behavior. *Developmental Psychology, 46*, 1-17. doi: 10.1037/a0017886

McCartney, K. (1984). Effect of quality of day care environment on children's language development. *Developmental Psychology, 20*(2), 244. doi: 10.1037/0012-1649.20.2.244

Minervino, J. (2014). Lessons from research and the classroom. Retrieved from [https://docs.gatesfoundation.org/documents/Lessons%20from%20Research%20and%20the%20Classroom\\_September%202014.pdf](https://docs.gatesfoundation.org/documents/Lessons%20from%20Research%20and%20the%20Classroom_September%202014.pdf) on March 28, 2018.

National Academies of Sciences, Engineering, and Medicine. (2018). *Transforming the Financing of Early Care and Education*. Washington, D.C.: The National Academies Press. Retrieved from <https://doi.org/10.17226/24984> on March 28, 2018.

National Association for the Education of Young Children. (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8: A position statement of the National Association for the Education of Young Children*. Washington, DC: Author. Retrieved from <http://www.naeyc.org/files/naeyc/file/positions/position%20statement%20Web.pdf> on March 28, 2018.

National Research Council. (2008). *Early childhood assessment: Why, what, and how*. Committee on Developmental Outcomes and Assessments for Young Children, C. E. Snow, & S. B. Van Hemel (Eds.). Board on Children, Youth, and Families, Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, D.C.: National Academies Press.

NICHD Early Child Care Research Network. (1997). Familial factors associated with characteristics of nonmaternal care for infants. *Journal of Marriage and the Family, 59*, 389-408. doi: 10.2307/353478

NICHD Early Child Care Research Network. (1999). Child outcomes when child-care center classes meet recommended standards for quality. *American Journal of Public Health, 89*(7), 1072-1077. doi: 10.2105/AJPH.89.7.1072

NICHD Early Child Care Research Network. (2002a). Child-care structure → process → outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science, 13*(3), 199-206. doi: 10.1111/1467-9280.00438

NICHD Early Child Care Research Network. (2002b). Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Educational Research Journal, 39*, 133-164. doi: 10.3102/00028312039001133

Office of Head Start (2018). *History of Head Start*. Retrieved from <https://www.acf.hhs.gov/ohs/about/history-of-head-start> on March 30, 2018.

Organisation for Economic Co-operation and Development (2018). Early Childhood Education and Care. Retrieved from <http://www.oecd.org/els/family/database.htm> on March 28, 2018.

Pace, A., Alper, R., Burchinal, M., Golinkoff, R., & Hirsh-Pasek, K. (2017). *Measuring success: Within and cross-domain predictors of academic and social trajectories in elementary school*. Paper presented at the Biennial Meeting of the Society for Research in Child Development. Austin, TX.

Peisner-Feinberg, E. S., & Burchinal, M. R. (1997). Relations between preschool children's child-care experiences and concurrent development: The cost, quality, and outcomes study. *Merrill-Palmer Quarterly, 43*(3), 451-477.

## CHAPTER 1 WHAT DOES RESEARCH TELL US ABOUT ECE PROGRAMS?

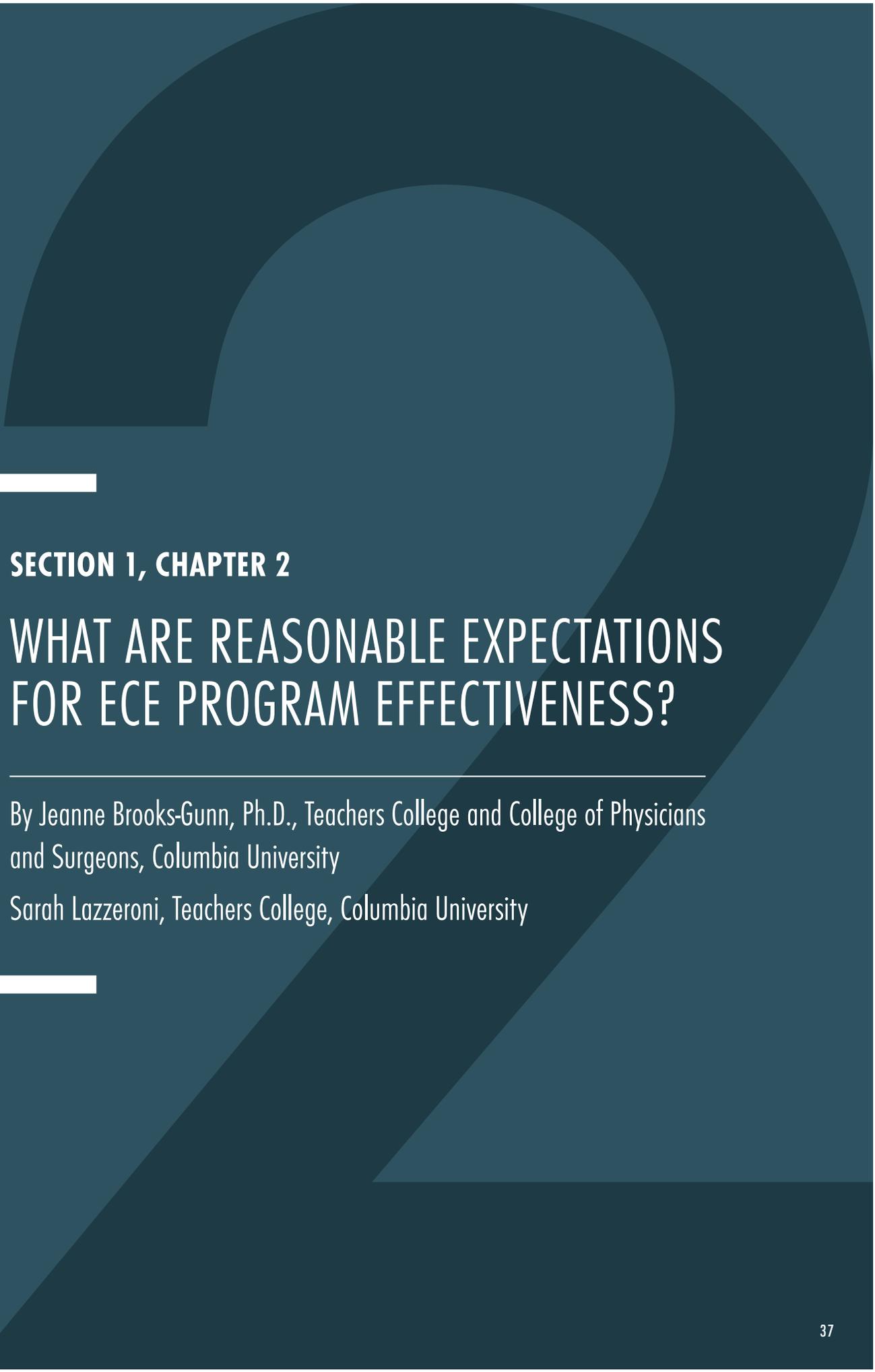
- Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Yazejian, N. (2001). The relation of preschool child care quality to children's cognitive and social developmental trajectories through second grade. *Child Development, 72*(5), 1534-1553. . doi: 10.1111/1467-8624.00364.
- 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*. Retrieved from [https://www.brookings.edu/wp-content/uploads/2017/04/consensus-statement\\_final.pdf](https://www.brookings.edu/wp-content/uploads/2017/04/consensus-statement_final.pdf) on March 28, 2018.
- Phillipsen, L., Burchinal, M., Howes, C., & Cryer, D. (1997). The prediction of process quality from structural features of child care. *Early Childhood Research Quarterly, 12*(3), 281-303. doi: 10.1016/S0885-2006(97)90004-1
- Pianta, R., Hamre, B., Downer, J., Burchinal, M., Williford, A., LoCasale-Crouch, J., ... Scott-Little, C. (2017). Early childhood professional development: Coaching and coursework effects on indicators of children's school readiness. *Early Education and Development, 28*, 956-975. <https://doi.org/10.1080/10409289.2017.1319783>
- Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R. M., Early, D. M., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Prediction of observed classroom quality and teacher-child interactions. *Applied Developmental Science, 9*(3), 144-159. doi: 10.1207/s1532480xads0903\_2.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (Pre-K)*. Baltimore, MD: Paul H. Brookes.
- Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M. J. (2010). Effects of an early literacy professional development intervention on Head Start teachers and children. *Journal of Educational Psychology, 102*, 299-312. doi: 10.1037/a0017763
- Puma, M., Bell, S., Cook, R., & Heid, C. (2010). *Head Start Impact Study: Final Report*. Washington, D.C.: U.S. Department of Health and Human Services, Administration for Children and Families. Retrieved from <http://www.eric.ed.gov/PDFS/ED507845.pdf> on March 28, 2018.
- Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., ... & Downer, J. (2012). Third Grade Follow-Up to the Head Start Impact Study: Final Report. OPRE Report 2012-45. Administration for Children & Families.
- Putnam, R.D., Frederick, C.B., Snellman, K. (2012). *Growing class gaps in social connectedness among American youth*. Cambridge: Harvard Kennedy School of Government.
- Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist, 53*, 109-120. doi: 10.1037/0003-066X.53.2.109
- Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Smallwood, K., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly, 23*, 10-26. doi: 10.1016/j.ecresq.2007.09.001
- Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Bub, K., & Pressler, E. (2011). CSRPs impact on low-income preschoolers' pre-academic skills: self-regulation as a mediating mechanism. *Child development, 82*(1), 362-378.
- Reynolds, A. J., Magnuson, K. A., & Ou, S. R. (2010). Preschool-to-third grade programs and practices: A review of research. *Children and Youth Services Review, 3*, 1121-1131. doi: 10.1016/j.childyouth.2009.10.017
- Ritchie, S., Weiser, B, Kraft-Sayre, M. & Howes, C. (2001). *Emergent academics snapshot scale* (Unpublished instrument). Los Angeles: University of California.
- Sabol, T. J., & Pianta, R. C. (2015). Validating Virginia's quality rating and improvement system among state-funded pre-kindergarten programs. *Early Childhood Research Quarterly, 30*, 183-198. <https://doi.org/10.1016/j.ecresq.2014.03.004>
- Schick, A. R. (2014). Home-school literacy experiences of Latino preschoolers: Does continuity predict positive child outcomes? *Journal of Applied Developmental Psychology, 35*(4), 370-380. doi: <https://doi.org/10.1016/j.appdev.2014.05.006>

- Snow, C. E., & Oh, S. S. (2010). Assessment in early literacy research. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research*. (Vol. 3, pp. 375-395). New York, NY: Guilford Press
- Hong, S. L. S., Howes, C., Marcella, J., Zucker, E., & Huang, Y. (2015). Quality Rating and Improvement Systems: Validation of a local implementation in LA County and children's school-readiness. *Early Childhood Research Quarterly*, 30, 227-240. <https://doi.org/10.1016/j.ecresq.2014.05.001>
- Spivak, A. & Farran, D. (2016). Predicting first graders' social competence from their preschool classroom interpersonal context. *Early Education and Development*. doi: 10.1080/10409289.2016.1138825
- Sprachman, S., Caspe, M., Atkins-Burnett, & Kennen, B. (2009). *Language Interaction Snapshot (LISn) Coding Manual*. Princeton, NJ: Mathematica Policy Research, Inc.
- Stipek, D., Franke, M., Clements, D., Farran, D., & Coburn, C. (2017). PK-3: What does it mean for Instruction? *SRCD Social Policy Report*, 30(2). ISSN 1075-7031, [www.srcc.org/publications/social-policy-report](http://www.srcc.org/publications/social-policy-report)
- Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2012). Pre-school quality and educational outcomes at age 11: Low quality has little benefit. *Journal of Early Childhood Research*, 9, 109-124. doi: 10.1177/1476718X10387900
- Thornburg, K. R., Mayfield, W. A., Hawks, J. S., & Fuger, K. L. (2009). *The Missouri Quality Rating System School Readiness Study: Executive summary*. Kansas City, MO: Center for Family Policy & Research University of Missouri and the Institute for Human Development University of Missouri.
- Tout, K., Cleveland, J., Li, W., Starr, R., Soli, M., & Bultnick, E. (2016). Parent Aware: Minnesota's quality rating and improvement system: Initial validation report. *Child Trends*, Bethesda, MD. <https://doi.org/10.1111/mono.12241>
- Vandell, D. L., Belsky, U., Burchinal, M., Steinberg, L., Vandergift, N., & NICHD ECCRN (2010). Do effects of early child care extend to age 15 years? Results from the NICHD Study of Early Child Care and Youth Development. *Child Development*, 81, 737-756. doi:10.1111/j.1467-8624.2010.01431.x
- Votruba-Drzal, E., Coley, R. L., & Chase-Lansdale, P. L. (2004). Child care and low-income children's development: Direct and moderated effects. *Child Development*, 75(1), 296-312. doi: 10.1111/j.1467-8624.2004.00670.x
- Vygotsky, D. H. (2001). *Pedagogy*. London and New York: RoutledgeFalmer, 69-95.
- Waldfogel, J., Han, W., & Brooks-Gunn, J. (2002). The effects of early maternal employment on child cognitive development. *Demography*, 39, 369-392. doi: 10.1353/dem.2002.0021
- Wasik, B. A., & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development. *Journal of Educational Psychology*, 103, 455-469. doi: 10.1037/a0023067
- Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly*, 28(2), 199-209. doi: 10.1016/j.ecresq.2012.12.002
- Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, 84, 2112-2130. doi: 10.1111/cdev.12099
- Yazejian, N., Bryant, D. M., Hans, S., Horm, D., St Clair, L., File, N., & Burchinal, M. (2017). Child and Parenting Outcomes After 1 Year of Educare. *Child Development*, 88(5), 1671-1688. <https://doi.org/10.1111/cdev.12688>
- 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. doi: 10.1037/a0038785.

## CHAPTER 1 WHAT DOES RESEARCH TELL US ABOUT ECE PROGRAMS?

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*. Ann Arbor, MI: Society for Research in Child Development.

Zellman, G. L., Perlman, M., Le, V.-N., & Setodji, C. M. (2008). *Assessing the validity of the Qualistar early learning quality rating and improvement system as a tool for improving child-care quality*. Santa Monica, CA: Rand Corporation.



**SECTION 1, CHAPTER 2**

**WHAT ARE REASONABLE EXPECTATIONS  
FOR ECE PROGRAM EFFECTIVENESS?**

---

By Jeanne Brooks-Gunn, Ph.D., Teachers College and College of Physicians  
and Surgeons, Columbia University

Sarah Lazzeroni, Teachers College, Columbia University



Early childhood education (ECE) programs in the U.S. have a long and rich history, as well as a robust evaluation literature. In fact, more well-designed evaluations have been conducted for ECE programs than for elementary or high school programs. Other chapters in this volume consider what we know about the quality of early childhood programs and child outcomes (Burchinal & Farran, Ch. 1), about instructional practices contributing to ECE quality (Farran, Ch. 4), and about how teacher training and professional development influence program quality (Pianta & Hamre, Ch. 5).

## EVALUATING ECE EFFECTIVENESS

---

This chapter examines a slightly different but related topic: What are reasonable expectations for ECE program effectiveness? The overlap is evident in that asking about expectations raises questions about what is reasonable today given the state of ECE quality, as well as the variability in quality. In general, ECE program impacts are expected to be small-to-medium, but not large.<sup>1</sup> Our estimates are based on the current ECE evaluation literature (Elango, Garcia, Heckman, & Hojman, 2015; Love, Chazan-Cohen, Raikes, Brooks-Gunn, 2013; Marietta, 2010; Phillips, Gormley, & Anderson, 2016; Weiland & Yoshikawa, 2013; Yoshikawa, Weiland, Brooks-Gunn, 2016). We offer general, research-based estimates for ECE program effectiveness. We should see modest program effects for four-year-olds whose teachers receive continuous professional development, a BA or additional training, adequate wages, and training on well-defined curricula. Additionally, all ECE programs should offer full-day programming and strive for relatively low teacher turnover. Some programs should be expected to enhance child school readiness by at least one-sixth to one-third of a standard deviation (more on this metric below). These effects would be found in traditional evaluations (randomization to treatment or control); they would be most likely in communities that do not have preschool slots for all four-year-olds (i.e., where a significant proportion of children are being cared for by kith and kin or where there is an age-based cutoff for enrollment).

This effect size is most likely to be seen in measures of language, literacy, mathematics, cognition, and perhaps, executive function (EF), which encompasses attention, memory, and inhibition. Significant effects are not likely for general health or health care, as the vast majority of four-year-olds are relatively healthy and receive health care. By contrast, if ECE programs offer referrals to or are linked with dental care, we are likely to see effects (since most four-year-olds, especially those who are poor or near poor, don't receive dental care). We can't be sure whether ECE increases receipt of services for special needs, as two opposing counterfactuals exist. That is, if ECE improves language, literacy, and cognition, then the proportion of children classified as developmentally delayed would decrease; at the same time, ECE program staff are likely to identify children who could benefit from Individuals with

---

<sup>1</sup> Large effects are defined as differences of .40 or more of a standard deviation between a group receiving an ECE program and a group not receiving the program (control group), where each child is ideally put into one of the two groups by random assignment. Medium effects are defined as .25 to .40 of a standard deviation. A small effect would be between .15 and .25 of a standard deviation (yet statistically significant). For the purposes of this chapter, our expected range is between one-sixth and one-third of a standard deviation.

Disabilities Education Act (IDEA) services, while kith-and-kin caregivers are much less likely to have the knowledge or access to do so (Love, Chazan-Cohen, Raikes, & Brooks-Gunn, 2013). In any case, ECE aims to provide needed services to each child, which would favor the second outcome (providing IDEA services).

We chose one-third of a standard deviation or more based on the best ECE evaluation results to date; not all program evaluations achieve this, for a variety of reasons (Yoshikawa, Weiland, & Brooks-Gunn, 2016). Having a robust effect size is also important given the expected reduction in effect sizes throughout the elementary school years. Without additional services or improvements to early elementary school, the effect of ECE will fall to one-half of its initial size by the end of third or fourth grade. Therefore, an effect size of one-half will become one-quarter and an effect size of one-third will become one-sixth. Effect sizes that are lower than one-third are very unlikely to be sustained into the late elementary school years.

It is likely that we would see smaller declines if changes, some of which we list below, were made in early elementary school. Without such changes, sustained ECE effects will be very modest or not present at all. Sixteen years ago, one of us wrote an article titled “Do You Believe in Magic?,” with a thesis that no matter how wonderful a preschool program might be, one year of even the highest-quality services is not enough for children to succeed (Brooks-Gunn, 2003). Improvements must be made in the quality and often the quantity of education at both the preschool and elementary school levels (not to mention middle and high school, but that is beyond the scope of this chapter). More time in education settings may also be necessary (for example, full-day pre-K and kindergarten and after-school and summer programs during elementary school).

Asking about reasonable expectations is especially important because almost three-quarters of adults are in favor of preschool programs (Jones, 2014). Most people appreciate the idea that an early start is one of the most effective approaches to helping children learn. In this sense, developmental psychologists and early childhood educators have been wildly successful. A few benefit-cost analyses—underscoring the message that earlier is better—have cemented this belief. Economists James Heckman and Lynn Karoly have provided compelling evidence of long-term effects (Cannon, Kilburn, Karoly, Mattox, Muchow, & Buenaventura, 2017; Heckman, 2006; Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010; Karoly, 2016). But underneath all the kudos lies a concern about what we should really expect from a preschool program in terms of children’s later well-being. Our success as educators and social scientists in communicating that an early start matters may have some unintended consequences. That is, expectations may outstrip results. Today’s ECE programs, even those showing short-term effects of one-third of a standard deviation, are unlikely to generate a 14:1 or even a 7:1 benefit-cost ratio, as the Perry Preschool Program did (Heckman, 2006). We believe that a more reasonable goal would be a 1.5 to 1 or 2 to 1 ratio of benefits to costs<sup>2</sup> (Karoly, 2016; Kilburn & Karoly, 2008; Steuerle, & Jackson, 2016).

Policy scholars debate (a) what effect sizes mean in terms of school achievement, (b) how large effect sizes need to be to translate into long-term indicators of success, (c) which ECE programs can deliver changes large enough to make a difference later on, and (d) whether our expectations for large effects are reasonable. We examine how ECE is defined, what types of evaluation are appropriate, how effect sizes are measured, what child outcomes are typically examined and what the results say (with a focus on differential effects), and what the implications are for pre-K to third grade education (Brooks-Gunn, 2003; Camilli, Vargas, Ryan, & Barnett, 2010; Duncan & Magnuson, 2013; Garces, Thomas, & Currie, 2002; Gormley & Gayer, 2005; Hill, Gormley, & Adelstein, 2015; Love, Chazan-Cohen, Raikes, & Brooks-Gunn, 2013; Reynolds, Magnuson, & Ou, 2010; Yoshikawa et al., 2013).

## DEFINING ECE PROGRAMS

---

In this chapter, early childhood education refers to programs that provide center-based education to children from one to five years of age. Center-based programs for children under one year, although they exist (the most notable being the Abecedarian Programs, Early Head Start programs, and the current Educare programs) (Yazejian, Bryant, Hans, Horm, St. Clair, File, & Burchinal, 2017), serve only a very small fraction of infants, given both the high cost of care in the first year of life and parental preferences. At five years old, most U.S. children enter kindergarten or at least become eligible for kindergarten. Currently, the vast majority of four-year-olds attend preschool, and the number of three-year-olds in preschool is rapidly rising: about 60% of four-year-olds (Rathbun, Zhang, & Snyder, 2016) and 43% of three-year-olds (Weiland & Yoshikawa, 2013) are enrolled in preschool, according to recent estimates (Yoshikawa, Weiland, & Brooks-Gunn, 2016). One- and two-year-olds are much less likely to attend preschool. Therefore, we focus on four-year-olds and, to a lesser extent, three-year-olds. (Most evaluations focus on four-year-olds, although they are beginning to include more three-year-olds, who are receiving ECE in increasing numbers.)

ECE programs have many goals. The primary goal is to envelop children in a learning milieu that provides opportunities to master age-appropriate social, emotional, linguistic, physical, and cognitive skills. A closely related focus is the relatively low levels of school readiness among some groups. Children whose parents have low education, low income, and/or are from minority ethnic groups are, on average, likely to enter kindergarten with lower skills than children from other backgrounds (Duncan & Magnuson, 2005; Reardon & Portilla, 2016). They are also less likely to receive high levels of learning stimulation at home (Brooks-Gunn, Markman-Pithers, & Rouse, 2016; Hoff, 2006; Hoff, 2012; Kalil, Ziol-Guest, Ryan, & Markowitz, 2016; Votruba-Drzal, 2003), in large part because of

---

<sup>2</sup> Sibling and county comparisons have been used to follow children into adulthood, in order to look at long-term sustained effects of ECE. A handful of the small-program evaluations have also done so (Abecedarian Project and the Perry Preschool Program) (Belfield, Nores, Barnett, & Schweinhart, 2006; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010; Hill, Gormley, & Adelstein, 2015). The estimates of effect sizes from these two programs are frequently cited by ECE policymakers as well as by politicians (one mention being made by President Obama in a State of the Union address). Although impressive, these benefit-cost estimates are based on fewer than 150 individuals who were born in the 1960s and 1970s.

their parents' education, income, and/or cultural beliefs about parenting. Many ECE programs have been designed to enhance poor children's school readiness, Head Start being the most salient. Sometimes the focus is on the gaps or discrepancies between more and less disadvantaged children. However, these terms do not address the goal of raising skills in one group (gaps could also be closed by reducing skill development in more advantaged groups).

Another goal is to help children who speak a language other than English at home to become proficient in both English and their native language. Policy scholars disagree about whether education for English language learners (ELLs) should focus more on helping students become bilingual or on helping students become proficient in English as quickly and efficiently as possible (Barrow & Markman-Pithers, 2016). Depending on which objective they emphasize, educational programs for young ELLs are generally divided between programs that are taught in both English and another language and programs that are taught solely in English (Barrow & Markman-Pithers, 2016).

**Higher labor force participation among women and more work hours have led to a need for safe, affordable, and educational child care, yet such care is not available to many... Our so-called polyglot system of early care and education is not conducive to supporting working parents.**

Last, and often overlooked, is the need for quality care for young children whose parents work. The proportion of working mothers with children age five and younger is at an all-time high in the U.S. (Bureau of Labor Statistics, 2016a; Bureau of Labor Statistics, 2016b; Wen, Hetzner, Brooks-Gunn, 2019). About 70% of all mothers with children under 18 are in the labor force, including 64% of mothers with children between the ages of one and five years (Bureau of Labor Statistics, 2016a; Bureau of Labor Statistics, 2016b). Many mothers in the U.S. also return to work quite soon after giving birth—almost 60% are back at work within nine months, 26% within 2 months, and 7% within one month (Wen, Hetzner, & Brooks-Gunn, 2019). Working hours have also increased, by 35% in single-parent households with children under age 18 and by 16% in two-parent households with children under age 18.

Higher labor force participation among women and more work hours have led to a need for safe, affordable, and educational child care, yet such care is not available to many (Chaudry, Morrissey, Weiland, & Yoshikawa 2017). Our so-called polyglot system of early care and education is not conducive to supporting working parents (Chaudry et al., 2017).

Although definitions vary, many use the term pre-K to refer to all early childhood educational programs (Brooks-Gunn, Markman-Pithers, & Rouse, 2016). Four categories of programs can be identified, depending on who administers the program and how it is funded. (Sometimes these lines are blurred since programs may be funded by more than one source and may be subject to multiple administrative rules; for example, see New York City's Pre-K for All program [Reid, Melvin, Kagan, & Brooks-Gunn, 2019]).

1. State or city pre-K programs are, for the most part, overseen by state or city education departments; they are often universal, although some may be targeted to low-income children (Friedman-Krauss, Barnett, Weisenfeld, Kasmin, DiCrecchio, & Horowitz, 2018).
2. Federally funded programs include Head Start and its younger sibling, Early Head Start. The U.S. Department of Health and Human Services administers these programs, which are targeted to families with income below the federal poverty threshold (with 10% of the Head Start children having special needs, as a mandated set-aside) (Elango, Garcia, Heckman, & Hojman, 2015).
3. Community programs include a panoply of not-for-profit programs. They may be subsidized by community organizations or by the Child Care Development Block Grant program, in which federal money is passed on to the states to subsidize child care costs for low-income working parents (Matthews, Schulman, Vogtman, Johnson-Staub, & Blank, 2015).
4. For-profit early childhood programs have not been studied extensively, although the few observations available suggest that their overall quality is lower than that of the other three categories (Burchinal, Nelson, Carlson, & Brooks-Gunn, 2008; Rathbun, Zhang, & Snyder, 2016).

In the 1960s through the 1980s, ECE programs were developed mostly for children whose parents had low incomes and/or low education. Children from such families were observed to be less prepared for kindergarten (academically and socially) than children from more advantaged backgrounds. In fact, gaps in language skills are seen as early as age two, and perhaps even earlier (Fryer & Levitt, 2013; Klebanov, Brooks-Gunn, McCarton, & McCormick, 1998). It was thought that children from educationally and economically disadvantaged households received fewer opportunities—in their families, neighborhoods, and child care settings—to develop early skills that predict literacy and numeracy (Blau, 2003; Johnson, Martin, & Brooks-Gunn, 2013; Noble, Houston, Brito, Bartsch, Kan, Kuperman, Akshoomoff, et al., 2015). Families who have low incomes or live in low-income neighborhoods are also constrained in their child care choices, due to both income and availability. These ECE programs were premised on the idea that an educationally oriented preschool would provide experiences that would reduce the gaps between economically disadvantaged kindergarteners and their more advantaged peers. Hence the term “Head Start,” the goal of which was to level the playing field by enhancing the skills of poor preschoolers. Consequently, programs from this era targeted children from low-income backgrounds. Thus, almost all the program evaluations through the last century involved children from low-income families. Our knowledge about program efficacy, especially long-term efficacy, is based on poor and, to a lesser extent, minority children. As more universal state and local pre-K programs have been implemented, we’ve seen debates arise about whether programs are equally effective for children from more advantaged families.

## CONSIDERING EVALUATION DESIGNS

---

### ► Commonly used designs

About 80% of the evaluations of ECE programs focus on four-year-olds (Camilli, Vargas, Ryan, & Barnett, 2010; Yoshikawa, Weiland, & Brooks-Gunn, 2016). Almost all the evaluations have been based on random assignment to a treatment or a control group. (A few well-known evaluations were not experimental—children were not randomized, and no data were collected prior to the treatment. The Chicago Parent-Child Program is the notable example [Reynolds, Temple, Robertson, & Mann, 2001].) These traditional evaluations are useful because they compare two equivalent groups of randomly assigned children. Therefore, any effects are unlikely to be due to unobserved differences between the two groups.

A few other designs have been used to evaluate ECE programs. One is based on sibling comparisons (looking at adolescent or adult outcomes of siblings who did and did not go to Head Start, for example) (Currie & Thomas, 1995), based on the premise that such comparisons control for family differences to a large extent. A few clever comparisons have employed variation in how programs were rolled out in a set of counties that were similar in poverty status, some of which received funding and technical assistance to open Head Start centers and some of which did not; this approach is a variant of the regression discontinuity design (Ludwig & Miller, 2007).

But such designs have limitations. Since parents voluntarily choose to send their children to ECE programs, the sample does not include families whose parents are unaware of a program, are distrustful of sending their children to a program, have few ECE programs available in their neighborhoods, do not speak English, or are concerned about immigration or child welfare scrutiny, to name a few of the reasons parents don't send their children to ECE programs. Consequently, we don't know how well an intervention may fare with all children of a specific age group. (Although citywide universal pre-K programs alleviate this concern to some extent, even in these circumstances, not all children are served.) And until recently, evaluations have focused on relatively small programs, offered in either just one site or in just a handful of sites. The national Head Start Impact Study (begun in 2002, even though Head Start itself began in 1965), which used a waitlist design, was the first to look at treatment and control children in hundreds of Head Start centers.

Evaluations of small programs are influenced by the community in which they are conducted. From an evaluation perspective, the biggest concern is the availability and quality of other ECE programs. If most children in a control group are likely to attend a different ECE program, then the effect sizes will be smaller than in situations where children in a control group do not attend an ECE program (Zhai, Brooks-Gunn, & Waldfogel, 2011). The other design that has been used to evaluate ECE programs is regression discontinuity, typically comparing children whose birthdays are near the mandated age cutoff for pre-K on either side. That is, children who receive ECE because their

birthdays are just before the age cutoff are compared to those who do not receive ECE because their birthdays fall right after it. Boston and Tulsa have used this evaluation design (Gormley & Gayer, 2005; Yoshikawa et al., 2013).

### ► The Counterfactual

The sibling and county comparisons also suffer from being based on ECE conditions almost 50 years ago. The sibling comparison analyses have tapped the Panel Study of Income Dynamics and the National Longitudinal Study of Youth—Child Supplement, which began in the late 1960s or the 1970s. The county comparison analyses were based on the first Head Start programs from the 1960s. Also, these studies focused on Head Start, which offers early childhood education only for children whose family incomes are at or below the poverty threshold. At the time, families with low income usually had no other options (few other programs were available in low-income neighborhoods, and even when other programs were available, families were often unable to afford them). Therefore, children who were not in Head Start were unlikely to be in other preschool programs or were in programs for only a few hours a day (see the ETS Head Start Evaluation from the 1970s as an example) (Lee, Brooks-Gunn, & Schnur, 1988; Lee, Brooks-Gunn, Schnur, & Liaw, 1990).

Today, children from low-income families have access not only to Head Start but also, in many cities and states, to universal pre-K programs, often run by or in collaboration with a department of education. Other partially subsidized programs also exist (some funded through the Child Care Development Block Grant). The two best-known, small-scale evaluations, the Perry Preschool and Abecedarian projects, also were initiated in the 1960s and 70s and also targeted poor children; very few of the children in the control groups received any other preschool experiences (Belfield, Nores, Barnett, & Schweinhart, 2006; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010; Hill, Gormley, & Adelstein, 2015).

All of this suggests that the counterfactual for treatment today is different from what it was previously. If children in control groups are enrolled in other preschool programs, the counterfactual is no longer preschool versus no preschool; it is a particular program (Head Start, universal pre-K) versus whatever other programs exist in a particular community. The heterogeneity within the control group vis-à-vis preschool experiences is important to quantify, and several nonexperimental analyses have been conducted to address it. Our group has done analyses with the Infant Health and Development Program (IHDP), the Head Start Impact Study, the Fragile Families and Child Wellbeing Study, and the Early Childhood Longitudinal Study—Kindergarten Cohort and Birth Cohort (Hill, Waldfogel, & Brooks-Gunn, 2002; Hill, Brooks-Gunn, & Waldfogel, 2003; Lee, Zhai, Brooks-Gunn, Han, & Waldfogel, 2014; Lee, Zhai, Han, Brooks-Gunn, & Waldfogel, 2013; Lee, Brooks-Gunn, Han, Waldfogel, & Zhai, 2014; Lee, Han, Waldfogel, & Brooks-Gunn, 2018). In all cases, we find the largest effects of Head Start, pre-K, or Learning Games (IHDP) occur in comparisons with children who received only parental or relative care, as well as in comparisons with home-based family care and home-based care with a nonrelative. These comparisons are

more akin to the analyses from the 1960s and 70s. Such findings—and their consistency across data sets—suggest that the effect sizes seen in the past are unlikely in current evaluations because more children in a control group are receiving some sort of preschool. Interestingly, comparisons of children in preschool or Head Start against children in kith-and-kin care show effect sizes in the modest range. Comparisons to children receiving other preschool do not.

**Consequently, it may be unreasonable to expect effect sizes today that are similar to those in the past if most children are now receiving some ECE at three and four years of age. This does not mean that preschool is ineffective.**

These findings have at least two implications. First, some preschool is better for children than none (even if quality differs across programs), as researchers have demonstrated in nationally representative longitudinal studies (Duncan & Magnuson, 2005; Duncan & Magnuson, 2013; Lee, Brooks-Gunn, Schnur, & Liaw, 1990; Lee, Zhai, Brooks-Gunn, Han, & Waldfogel, 2014; Yoshikawa et al., 2013). Second, although specific programs that are believed to be of high quality are likely to be better than other programs presumably of lower quality, these differences will be smaller than what was seen in the past, given that the counterfactual is different (Duncan & Magnuson, 2013). Consequently, it may be unreasonable to expect effect sizes today that are similar to those in the

past if most children are now receiving some ECE at three and four years of age. This does not mean that preschool is ineffective. It just means that traditional evaluations of treatment and control will find smaller effect sizes, since most children in the control group are receiving some sort of preschool.

### ► **Alternative evaluation approaches**

The evaluation approaches discussed above are often considered superior to others, but they do have limitations, the most serious having to do with external validity, generalization, and take-up. Other approaches include using district-wide achievement test scores to examine cohorts before and after a district-wide intervention is initiated (see the example of Montgomery County discussed below). Another is to employ much more short-term, small-scale interventions to test a particular innovation before implementing it on a broad scale, or even before a traditional randomized trial to test for efficacy. An example of this approach has been outlined by Fisher et al. (2016) and Shonkoff & Fisher (2013).

Yet another approach is to forgo assessment of children altogether and, instead, focus on documenting changes made on quality indicators (see Burchinal & Farran and Pianta, this volume). Of course, such an approach is based on a strong premise—that quality is associated with child outcomes and that increasing the former improves the latter. (For example, if child outcomes are enhanced only when a certain level of quality is obtained [threshold effect] or if only children who initially experience a very low-quality program are affected [baseline effect], then just documenting quality increases cannot be assumed to result in more school readiness). Indeed, the ECE evaluation field is still struggling with the question of how much and what types of quality improvement actually make a difference.

## DESCRIBING EFFECT SIZES

---

### ► A definition of effect sizes

Evaluations typically report findings in terms of effect sizes as a function of the standard deviation; evaluation research often defines large effects as two-fifths to one-half of a standard deviation (with an assessment normed to have a mean of 100 and a standard deviation of 15, the treatment group would have a 6- to 8-point advantage at the end of an intervention compared to the control group) (Barnett, 2008).

### ► Effect sizes in everyday language

It is sometimes difficult for the public, policymakers, and educators to understand what an effect size means. For example, does an effect size of .40 on early indicators of achievement for low-income students mean they'll do better in elementary school, and how much better compared to high-income students? The same question, of course, could be asked for dual language learners or for minority students. Two approaches can help translate effect sizes into more concrete indicators. The first is to explain what might be seen in a classroom where low-income students' performance was one standard deviation below that of high-income students. As a heuristic, we are using the difference between students whose family incomes are in the bottom 10% and students whose family incomes are in the top 10% of the income distribution (Reardon, 2011). The following discussion is taken from Rock and Stenner (2005); they were comparing black and white students, not low-income and high-income students, but the general principle is the same. Based on a normal distribution (68% of scores will be within one standard deviation of the mean score, the difference between the peak of the distributions is one standard deviation, and the distributions for both groups are "normal"), the following estimates can be made:

*First, randomly selecting one black child and one white child and comparing their scores will show the white child exceeding the black child 76% of the time and the black child exceeding the white child 24% of the time. Second, 84% of white children will perform better than the average black child, while 16% of black children will perform better than the average white child. Third, if a class that is evenly divided by race is divided into two equal-sized groups based on ability, then black students will compose roughly 70%, and whites 30%, of the students in the lower performing group. Fourth, if a school district chooses only the top-scoring 5% of students for "gifted" courses, such classes will have thirteen times more whites than blacks. Fifth, assume that a school district's student body mimics the national racial distribution (17% black, 83% white and other). The district chooses the lowest-scoring 5% of all students for a special needs program. Although 17% of the district's children are black, 72% of the special needs students will be black (pp. 26-27).*

If low-income students benefit from an ECE program with an effect size of one-half of a standard deviation, then the difference between low- and high-income children would be reduced by one-half (assuming that the low-income students, rather than both groups, received the treatment, or that the low-income students responded twice as much to the treatment as did the high-income students). The corresponding changes in the differences between the hypothetical students in an example like the one above would be very large. Even benefits of one-third of a standard deviation would be considered large.

### ► Estimation of adult outcomes

Another approach is to take an effect size at the end of a preschool intervention and estimate the increase in the number of children graduating from high school or college or predict kindergarten achievement scores to high school achievement scores. Then the adolescent outcomes become the predictors for adult success (i.e., lifetime earnings). Brooks-Gunn, Magnuson, and Waldfogel (2009) used this estimation approach to see to what degree different effect sizes from preschool interventions are associated with gains in lifetime earnings. Card and Krueger (1996) used a similar procedure to estimate the long-term effects of reductions in elementary school class sizes, and Heckman et al. (2009) have done estimates using actual earnings data from the Perry Preschool Project. These estimates do not look at reducing the gap between groups of students, as the Rock and Stenner (2005) estimates do.

### ► Differential effectiveness for poor and nonpoor children

The example used here is based on one of the goals of ECE, which is to improve school readiness for disadvantaged children (whose parents are poor, have little education, are immigrants, do not speak English well, or are from minority backgrounds), targeting health and emotional, literacy, and cognitive skills. Some ECE programs are taking

**If all four-year-olds receive quality ECE, the differences between advantaged and disadvantaged students are likely to be smaller, unless large differential benefits are seen among groups.**

a different approach, targeting an entire school district. If all four-year-olds receive quality ECE, the differences between advantaged and disadvantaged students are likely to be smaller, unless large differential benefits are seen among groups. That is, both advantaged and disadvantaged children will benefit (a rising tide lifts all boats). Remember that until very recently, ECE program evaluations have concentrated on groups likely to have

lower rates of school readiness. Universal services may need to be evaluated differently, or at the very least, the possibility of not attenuating differences between groups needs to be explicitly addressed, and it is important to examine the specific mechanisms that lead to such differences. For example, in the Boston program, effects differ based on subgroup status—the program had higher effect sizes for low-income children than for higher-income children for numeracy, inhibitory control, and attention shifting (Weiland & Yoshikawa, 2013).

### ► Benefit-cost analyses

As another example, the varying estimates from benefit-cost analyses are confusing. Such estimates, of course, are based on myriad decisions (Steuerle & Jackson, 2016) on both the cost and the benefit sides of the equation. For example, benefit-cost estimates for the Perry Preschool Project range from 17:1 to 3:1, a huge range (and for the gender-linked estimates, the comparisons involve about 40 treatment and 40 control group boys and involve lower crime rates for the boys in the treatment group, meaning that the large benefit-to-cost ratios are based on about four fewer boys in the treatment group having been involved in a serious crime than in the control group) (Barnett, 1985; Belfield, Nores, Barnett, & Schweinhart, 2006; Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010). And the chances that a preschool program today will result in even a 3:1 savings (let alone a 17:1 savings) are likely to be small, given the counterfactual. It may be time for those of us in ECE to manage expectations by making it clear that benefit-cost ratios are likely to be no greater than 2:1.

## CHOOSING DEVELOPMENTAL OUTCOMES

---

What outcomes are preschool programs expected to influence, given that the goal is usually enhanced readiness for kindergarten and elementary school? School readiness is typically considered to encompass all facets of children's development—language and cognition, social and emotional development, physical growth and health, approaches to learning and persistence, enthusiasm, and motivation. Today, EF would be added as a separate facet, given its links to emotional and cognitive development (Raver & Blair, 2016). Educators and developmental psychologists may parse the domains a bit differently, yet they agree in looking at what they call the “whole child,” rather than at just academic achievement. At the same time, most preschool programs privilege some domains over others, with language and cognitive development—as reflected in achievement test scores—being the most desirable (and measured) outcome. Whether the implicit move away from the whole child approach is merited, given what we know about development and learning, is an open question. Most practitioners and evaluators are calling for less reliance on achievement test scores, and efforts to measure other domains continue. The relatively recent addition of EF outcomes to evaluations is a good example, as EF is thought to be central for learning and achievement. At the same time, emphasis on physical development and health is waning. We suspect that such changes are driven in part by the increase in state and local pre-K programs that are primarily administered through or with education departments rather than health and human services departments. Most educators subscribe to the belief that children need to be healthy to learn most effectively, but most programs don't emphasize health, per se.

All educational programs focus on language and cognition. Often these are defined in terms of achievement rather than developmental outcomes. Most curricula and teacher training emphasize literacy, numeracy, and science skills appropriate for each age group. Generic curricula are most often used in preschool programs, especially in Head Start (for example, Creative Curriculum). Preschool literacy curricula have the most extensive history (although, perhaps surprisingly, they have not been subject to evaluation; see Snow & Matthews, 2016). Specific numeracy

curricula have been developed and evaluated extensively (Building Blocks being the notable example), but science curricula for preschool children are less refined (Clements & Sarama, 2016). New approaches to enhancing EF are also being evaluated (Raver & Blair, 2016). In addition, some emerging curricula integrate learning across domains.

Head Start explicitly includes physical skills and health outcomes among its goals, whereas state and local pre-K are less likely to do so. Head Start funds services like occupational and physical therapy, and it offers health checkups and referrals for various forms of health care (like dental care). Such services are not the purview of schools, so they are less likely to be funded in state pre-K programs (although children who qualify for IDEA presumably would be referred for occupational and physical therapy, if indicated). When Head Start began, children from low-income families were unlikely to get regular health care; one of Head Start's successes in the past century was ensuring that high proportions of children obtained such services. Today, with more children covered by Medicaid and CHIP, differences in receiving health care between children in Head Start and not in Head Start are quite small (the exception being children whose parents are immigrants, who are less likely to receive health care than children whose parents are not immigrants). However, Head Start today does make a large difference in dental care, which many low-income children don't receive. Links to diagnostic and screening services may also increase the likelihood of receiving special education services through IDEA. Given Head Start's mandate to set aside slots for children with special needs, it is likely that Head Start serves proportionately more such children than do state and local pre-K (Reid, Melvin, Kagan, & Brooks-Gunn, 2019).

Interestingly, evaluations of ECE programs almost always include indicators of disabilities and individualized education plans. But we know little about whether teachers who have special needs children in their classrooms have received appropriate training or whether they provide specific or modified instruction for these students (Hebbeler & Spiker, 2016), let alone the additional services that children are receiving through IDEA. Few evaluations assess activities of daily living, a common measure in health surveys. Nor do they measure common health problems, such as asthma, which if not controlled is linked to school absence (Currie, 2005). Evaluations also measure more general indicators of health, such as weight for height (the concern being overweight and obesity, not underweight), nutritional intake (usually general measures), and exercise patterns. Whether programs actually emphasize such health behaviors is not known (Head Start does so, although very little is known about how much attention any individual program gives to health) (Lee, Zhai, Han, Brooks-Gunn, & Waldfogel, 2013).

Evaluations also often assess emotional development, most often in terms of aggression and inattention, as it is believed that disruptive behaviors impair the learning of individual children and in the classroom as a whole (Georges, Brooks-Gunn, & Malone, 2012; Duncan et al., 2007). We know less about how teachers actually manage such behaviors (and about how they are trained to do so) than about how teachers provide instruction in literacy and numeracy (Raver & Blair, 2016; Raver, Jones, Li-Grining, Metzger, Smallwood, & Sardin, 2009). Even so, reducing aggressive and inattentive behaviors is seen as an outcome of ECE programs. Likewise, what educators call "approaches to

learning,” or what psychologists term “motivation, enthusiasm, and persistence,” are often measured. As in the case of aggression, what teachers actually do to enhance motivation has not been studied very well.

One takeaway from this brief discussion of preschool outcomes is that links are often tenuous between expectations for children’s success or preparation for elementary school and what is known about curricula, teacher training, and even teacher behavior in the classroom. The notable exception is for literacy and numeracy achievement (Clements, & Sarama, 2016; Snow & Matthews, 2016). If we expect an ECE program to reduce aggression and inattention, enhance motivation and enthusiasm, promote healthy eating, increase EF, or decrease school absences due to illness, we will need to specify (and implement) classroom practices that explicitly target these outcomes.

**If we expect an ECE program to reduce aggression and inattention, enhance motivation and enthusiasm, promote healthy eating, increase EF, or decrease school absences due to illness, we will need to specify (and implement) classroom practices that explicitly target these outcomes.**

## EXAMINING ECE EFFECT SIZES

Preschool’s efficacy has been examined in over 120 evaluations (Brooks-Gunn, Markman-Pithers, & Rouse, 2016; Camilli, Vargas, Ryan, & Barnett, 2010; Yoshikawa, Weiland, & Brooks-Gunn, 2016). In general, evaluations report significant effects for four-year-olds. Recent evaluations show that preschool has positive effects in the short term on language, literacy, and math skills, with higher-quality programs showing the biggest effects (Yoshikawa, Weiland, & Brooks-Gunn, 2016). Some evidence suggests that preschool may have positive effects on socioemotional behaviors (e.g., decreased aggressive behavior), although the research in this area is not as definitive (Yoshikawa, Weiland, & Brooks-Gunn, 2016). But the range of effects is large. Even the early programs from the 1960s and 70s exhibited a range, although we usually emphasize the successful programs from that era (Brooks-Gunn & Hearn, 1982; Stipek, Franke, Clements, Farran, & Coburn, 2017). This state of affairs continues today; as examples, we have only to look at the Head Start Impact Study results (small effects at the end of the program with few effects sustained into elementary school) (U.S. Department of Health and Human Services, 2010) and the Tulsa Head Start results (large and sustained effects seen through middle school) (Gormley & Gayer, 2005; Phillips, Gormley, & Anderson, 2016). How do we interpret such disparate findings? Other authors in this volume focus on program quality and implementation (the two are difficult to separate), curricula, and teacher training and oversight. The composition of students in a classroom also matters (via a process economists often call heterogeneity of effects). Some groups, such as students with developmental disabilities and dual language learners, have not received enough attention regarding effective ways of teaching and including them in classrooms (Barrow & Markman-Pithers, 2016; Hebbeler & Spiker, 2016).

The evaluation literature is replete with examples of differential effectiveness across subgroups within a center, across types of centers, and even across centers under the same auspices. Such variation makes it difficult to say

what expectations may be reasonable for outcomes in different programs. We provide a few examples, making comparisons within and across centers.

### ► **Comparisons within centers**

Within centers, comparisons have examined which subgroups benefit the most from ECE programs. Yoshikawa et al. (2013) looked at the effects of ECE programs on four sometimes overlapping subgroups: 1) poor and nonpoor children; 2) black, white, and Hispanic children; 3) dual language learners and children of immigrants; and 4) children with special needs/disabilities. Gaps in school readiness based on income and race/ethnicity appear as early as age two, when children from nonpoor families and white children perform better on measures of literacy and cognitive skills (Brooks-Gunn, Markman-Pithers, & Rouse, 2016; Garces, Thomas, & Currie, 2002; Snow & Matthews, 2016). Preschool enrollment is lower for minority children and children from low-income families than for white children and children from higher-income families, possibly contributing to this gap (Brooks-Gunn, Smith, Klebanov, Duncan, & Lee, 2003; Reardon & Portilla, 2016; Yoshikawa, Weiland, & Brooks-Gunn, 2016). However, preschool's positive effects in terms of literacy, math, and social-emotional skills may be most effective for children living in or near poverty (Yoshikawa et al., 2013).

Most early evaluations have examined ECE programs' effects on black children but not on other minority groups, and therefore they can't give full insight into differential ECE program effects based on race/ethnicity (Bassok, 2010; Yoshikawa et al., 2016). In response, recent evaluations of programs like Head Start, Tulsa Pre-K, and Boston Pre-K have made comparisons across racial groups. These programs showed positive effects for children of all racial/ethnic backgrounds, but the effects were highest for Hispanics at age three in Head Start and in both the Tulsa and Boston studies (Yoshikawa et al., 2016). Some studies found especially strong effects for minority children from low-income families (Love, Chazan-Cohen, Raikes, & Brooks-Gunn, 2013). Other studies found no racial differences for children living below the poverty line, but more benefits for black students than for whites or Hispanics among the nonpoor (Bassok, 2010).

Although research on ECE's effects on ELLs and children of immigrants is somewhat limited, some evidence suggests that ECE has positive effects on language development and cognitive skills for ELLs (Barrow & Markman-Pithers, 2016; Yoshikawa et al., 2016). Policy scholars debate whether language instruction should be conducted solely in English, or in a combination of English and children's first language, but Barrow and Markman-Pithers, (2016) find that the general quality of ECE programs may be more important than the language of instruction. Still, some evaluations show that dual language instruction does not hurt children's ability to learn English and may encourage bilingualism and even achievement overall (Hoff, 2012; Yoshikawa et al., 2016).

We also have few evaluations (especially randomized controlled trials) of ECE's impacts on children with disabilities (Hebbeler & Spiker, 2016). Head Start has shown positive effects on math and social-emotional skills for children

with disabilities, and Tulsa showed positive effects on their literacy skills (Gormley & Gayer, 2005; Yoshikawa et al., 2016). Some effective interventions include programs emphasizing language development and social-emotional development, which have been shown to be effective in promoting language/literacy skills and social skills, respectively (Raver & Blair, 2016). Additionally, specialized curricula and instructional strategies for children with disabilities have been shown to improve children's oral language, literacy, motor, and social skills (Hebbeler & Spiker, 2016). However, we need more evaluations of ECE's effectiveness for children with disabilities.

### ► Comparisons across Head Start centers

Comparisons are also being made across Head Start centers. Head Start has clear and extensive standards, which might lead us to expect that variation in impacts from site to site might be small. Yet such differences exist. In one analysis, inter-center variation was found for language and literacy but not for mathematics (U.S. Department of Health and Human Services, 2010). One possible explanation is that Head Start teachers generally are not doing much in the way of math instruction (see Clements, Sarama, & Germeroth, 2016, for evidence that in general, pre-K teachers are not spending much time on math and that when they do, they focus on simple math skills). Low math skills among students across the board would be evidence that such an explanation is correct. Head Start teachers do focus on language and literacy; the differences in outcomes suggest that some teachers are more effective than others. However, we need to know more about what exactly teachers are doing in literacy instruction (Snow & Matthews, 2016).

Another (nonexperimental) analysis from the Head Start Impact Study suggests that full-day programs had larger effects than half-day programs, which is not surprising (Yoshikawa, Weiland, & Brooks-Gunn, 2016; Yoshikawa et al., 2013). What is perhaps surprising is that teacher education (BA), teacher training (teaching license), and student-teacher ratios were not associated with inter-center program impacts (Yoshikawa, Weiland, & Brooks-Gunn, 2016). Still, a new analysis by Morris et al. (2018) suggests that Head Start's positive impacts are more variable than impacts shown in previous analyses, such as the U.S. Department of Health and Human Services' Head Start Impact Study from 2010. Morris et al. (2018) found that the effect sizes of Head Start on enrollment and exposure to high-quality care varied by site, with standard deviations of 21.4 percentage points (any center care), 22.3 percentage points (Head Start care), and 28.4 percentage points (nonrelative care). This variation may be due to differences in state regulations and implementation, as well as to variation in child characteristics (e.g., pretest scores and dual language learners).

### ► Comparisons across types of centers

Generally, children attending either Head Start, pre-K, or other center-based care performed better on academic-skill assessments than children in parental or relative care (Zhai, Waldfogel, & Brooks-Gunn, 2013), and recent studies have begun examining differences in the effects of different types of center-based programs. Children in Head Start performed better on reading and math assessments than children in parental care, pre-K, or other center-based care (ECLS-B data; Lee, Zhai, Brooks-Gunn, Han, & Waldfogel, 2014; Lee, Zhai, Han, Brooks-Gunn, & Waldfogel, 2013).

Additionally, children spent more hours in Head Start, on average, than children spent at non-Head Start centers (Lee, Zhai, Brooks-Gunn, Han, & Waldfogel, 2014). This increased exposure could be one of the mechanisms behind the finding that three- and four-year-olds attending Head Start fared better in classroom literacy and math instructional activities than children in non-Head Start centers (U.S. Department of Health and Human Services, 2010).

But other analyses conducted with data from the Fragile Families and Child Wellbeing Study showed that Head Start attendance was not significantly associated with cognitive gains when compared to attending pre-K or other center-based care (Zhai, Brooks-Gunn, & Waldfogel, 2011). Similarly, an analysis of Head Start Impact Study data found more substantial differences between children attending Head Start and children in parental or relative/nonrelative care than between children attending Head Start and children attending other center-based care (Zhai, Brooks-Gunn, & Waldfogel, 2014).

### ► Understanding the reduction in effect sizes in elementary school

Evaluations show that ECE programs have positive short-term effects. But multiple studies show that these effects fadeout (or decrease) by the third grade, with a decline of up to .03 per year in effect sizes for cognitive and test score outcomes (Camilli, Vargas, Ryan, & Barnett, 2010). Fadeout is also called the “convergence” or the “catch-up effect,” as the gap in achievement between children who attended (and benefited from) ECE programs and children who did not attend such programs decreases as the children get older (Yoshikawa et al., (2013). Eventually, children without any ECE perform as well as children who received ECE. However, receiving ECE is positively related to other long-term outcomes, such as higher earnings and a lower likelihood of criminal activity (Duncan & Magnuson, 2013; Karoly, 2016).

Across almost all experiments, effect sizes from ECE evaluations fall by one-half, on average, between the end of the program and the middle of elementary school. At the moment, this evidence is based almost exclusively on achievement test scores, although a few evaluations have reported a similar decline for aggressive behaviors and approaches to learning. Consequently, a reasonable expectation is that unless changes are made to K–3 education, sustained effects will be one-half the size of short-term effects.

*The possible reasons for this decline include:*

1. Children who did not receive ECE use kindergarten and first grade to catch up to their peers, mastering comparable skills later than children who received ECE (Duncan & Magnuson, 2013).
2. Early elementary school teachers may emphasize skills that children do not have (i.e., they direct teaching toward students with lower skills, including those who may not have had any preschool education) (Duncan & Magnuson, 2013).

3. Differences in curricular content and quality of instruction may also contribute to the fadeout of ECE's positive effects. Another potential cause is a lack of integration between preschool and elementary school curricula. Continuity between ECE and elementary school curricula is important for sustaining effects over time; some interventions, including one in Maryland's Montgomery County, have implemented such continuity (Brooks-Gunn, Markman-Pithers, & Rouse, 2016). When curricula are integrated, skills developed in ECE can be practiced and reinforced in elementary school.
4. In terms of instructional style, elementary schools may emphasize individualized learning less than preschools do. Preschool classes have lower adult-child ratios than elementary schools; preschool classes are often limited to 20 students, while elementary school classes often have 26 to 30 students (Pianta, Downer, & Hamre, 2016). In one study in Tennessee, smaller classes in elementary school were associated with better cognitive outcomes (Mosteller, 1995; Heckman, 2006). Individualized instruction has been shown to be most effective for learning outcomes (Clements & Sarama, 2016; Hebbeler & Spiker, 2016), and increased class sizes hinder teachers' ability to provide high-quality interactions with children. Moreover, low-income students are likely to attend elementary schools with larger class sizes, which are associated with lower achievement in general and may dilute preschool gains.
5. Instructional quality may also vary more in elementary school than in preschool, or quality may matter more for learning in elementary school (Pianta, Downer, & Hamre, 2016). For example, students from low-income backgrounds and students from racial/ethnic minority groups—for whom ECE was developed and who tend to benefit most from ECE—often receive low-quality instruction in elementary school (Burchinal, Howes, Pianta, Bryant, Early, Clifford, & Barbarin, 2008; Mashburn, Pianta, Hamre, Downer, Barbarin, Bryant, & Howes, 2008; Moiduddin, Aikens, Tarullo, West, & Xue, 2012). Students attending low-quality elementary schools cannot build on or sustain gains made in preschool, and the positive effects of preschool become less apparent. Low-income students who attended preschool may also be more likely to attend schools in communities where after-school programs, an extended school year, and other enrichment activities are not offered, making it difficult to sustain effects.
6. Elementary schools also tend to provide less support to parents than preschool programs do—especially ECE programs that primarily serve low-income families. For example, Early Head Start offers home visiting, referrals for health care, and parent education. Similarly, the Tulsa program offers parent education, health and vision screenings, and child care services. Such comprehensive supports have been shown to improve cognitive, academic, and health outcomes for children, but elementary schools don't often offer them (Phillips, Gormley, & Anderson, 2016). More comprehensive services for parents and families during elementary school might help sustain ECE's positive effects (Reynolds, Magnuson, & Ou, 2010), although little is known about the efficacy of such efforts (Magnuson & Schindler, 2016).

## PROMOTING SUCCESS: A MULTILEVEL MODEL

---

### ► Multiple changes in pre-K to 3: A model for ensuring success?

Almost all ECE evaluations have assessed individual children, typically those who received an intervention and those who did not via random assignment, waitlist, or eligible age for entrance into preschool. But some have used administrative data as well. One interesting approach is to analyze school- or district-wide data from standardized testing to look at differences in achievement levels. In this way, comparisons can be made across time to see whether an intervention implemented at the school or district level has increased mean scores or competency levels. Such a design is a variant of regression discontinuity.

Such a cohort comparison was used effectively in the county-level effort in Montgomery County, MD (Marietta, 2010). The school district staff, after examining the proportion of the district's high school seniors who were ready for college, set a goal of having 80% of a graduating class college-ready. Working backward, they defined their goals for classes of pre-K to third-grade children. They aimed to increase the percentage of children reading at grade level in the early grades. They then made a list of possible reforms that based on research were likely to prepare their young students to eventually be ready for college.

*The reforms were extensive, underlying the fact that no single change is likely to have large effects.*

*The county applied most of the changes recommended by early childhood educators. These included:*

1. pre-K for all four-year-olds,
2. full-day pre-K,
3. full-day kindergarten,
4. after-school programs,
5. summer programs,
6. curricula aligned across the early grades,
7. student-teacher ratios of only 15 to 1 from pre-K to third grade,
8. pre-K teachers having a BA and being certified in ECE,
9. earnings of pre-K teachers at parity with teachers in kindergarten to third grade,
10. English as a second language courses for parents, and
11. welcome packets and curricular guidebooks for parents of entering kindergartners.

This intensive and extensive set of reforms doubled the percentage of children reading at grade level by third grade, and this proportion was sustained through the later elementary school years (Marietta, 2010).

Most system-wide initiatives have not taken Montgomery County's approach to evaluation. And such initiatives have not coupled school-level data with individual-child data. Putting the two together might, for example, let us discover which subgroups of children are most likely to see an increase in the share of students reading at grade level, or which set of services are most likely to produce a higher proportion of competent readers.

## CONCLUSION

---

Programs that report sustained effects in elementary school and beyond typically have large effects at the end of an ECE program. Is it critical to have effect sizes of about one-sixth to one-third of a standard deviation at the end of a program to have any chance of seeing sustained effects? The evidence to date suggests that the answer is yes, absent changes in elementary school.

Therefore, we should try to amplify effect sizes in ECE programs, in the hope of improving both short- and long-term outcomes for children. Multiple steps could be taken to increase effect sizes in preschool. First, it is important to increase the dosage and duration of preschool. To increase dosage and duration, it is recommended that students attend preschool more days per year, and even that children attend two consecutive years of preschool. Additionally, there is some evidence that full-day programs are more effective than partial-day programs. Second, it is important to develop and implement more targeted and integrated curricula in preschool. Curricula should be developmentally appropriate and should aim to help children develop essential cognitive and social-emotional skills, as well as ensure that children have the necessary academic skills for elementary school. Moreover, preschool curricula and elementary school curricula should be integrated in an attempt to ensure continuity between the two programs. Third, to ensure the effective implementation of the targeted and integrated curricula, teachers need to be better trained. Fourth, programs should focus on teacher support and scaffolding of skills. Adequate support allows teachers to use structured, individualized teaching models that help children progressively build skills. Last, programs should target poor, minority, and immigrant children to narrow some of the early gaps in math and language literacies.

Although early childhood education programs like Tulsa's Head Start and Boston's Pre-K initiatives provide encouraging support for further investment in early childhood education, we should be specific in determining where and when to invest. Numerous studies have illustrated that it's important to increase young children's exposure to ECE while also working to ensure that quality is consistent across sites and types of programs. Further, the connection (in terms of curriculum, outcomes, and quality) between ECE and K-12 education should be strengthened to promote the maintenance of ECE gains. Policymakers should aim to use the lessons from previous evaluations to improve ECE programs in hopes of reducing achievement gaps and preparing young children for elementary school and beyond.

## References

---

- Barrow, L., & Markman-Pithers, L. (2016). Supporting young English learners in the United States. *Future of Children, 26*(2), 159-183.
- Barnett, W. S. (1985). Benefit-cost analysis of the Perry Preschool Program and its policy implications. *Educational Evaluation and Policy Analysis, 7*(4), 333-342.
- Barnett, W. S. (2008). *Preschool education and its lasting effects: Research and policy implications*. Boulder and Tempe, CO: Education and the Public Interest Center & Education Policy Research Unit.
- Bassok, D. (2010). Do black and Hispanic children benefit more from preschool? Understanding differences in preschool effects across racial groups. *Child Development, 81*(6), 1828-1845.
- Belfield, C., Nores, M., Barnett, W.S., & Schweinhart, L. (2006). The High/Scope Perry Preschool Program: Cost-Benefit Analysis Using Data from the Age-40 Follow-Up. *Journal of Human Resources 41*(1), 162-190.
- Blau, D. (2003). Childcare subsidy programs. In R. A. Moffitt (Ed.), *Means-tested transfer programs in the United States* (pp. 443-516). Chicago: University of Chicago Press.
- Brooks-Gunn, J. (2003). Do you believe in magic? What we can expect from early childhood intervention programs. *Social Policy Report, Society for Research in Child Development, 17*, 3-15.
- Brooks-Gunn, J., & Hearn, R. (1982). Early intervention and developmental dysfunction: Implications for pediatrics. *Advances in Pediatrics, 29*, 497-527.
- Brooks-Gunn, J., Magnuson, K., & Waldfogel, J. (2009). Long-run economic effects of early childhood programs on adult earnings. *Partnership for America's Economic Success Issue Report #12*, 1-34. Retrieved from [http://www.readynation.org/docs/researchproject\\_adultearnings\\_200903\\_brief.pdf](http://www.readynation.org/docs/researchproject_adultearnings_200903_brief.pdf)
- Brooks-Gunn, J., & Markman, L. (2005). The contribution of parenting to ethnic and racial gaps in school readiness. *Future of Children, 15*, 139-168.
- Brooks-Gunn, J., Markman-Pithers, L., & Rouse, C. E. (Eds.). (2016). Starting early: Education from prekindergarten to third grade (special issue). *Future of Children, 26*(2).
- Brooks-Gunn, J., Smith, J., Klebanov, P. K., Duncan, G. J., & Lee, K. (2003). The black-white test score gap in young children: Contributions of test and family characteristics. *Applied Developmental Science, 7*, 239-252.
- Bureau of Labor Statistics. (2016a). *Facts over time*. U.S. Department of Labor. Retrieved from [https://www.dol.gov/wb/stats/facts\\_over\\_time.htm](https://www.dol.gov/wb/stats/facts_over_time.htm)
- Bureau of Labor Statistics. (2016b). *Employment characteristics of families-2015*. U.S. Department of Labor. Retrieved from <https://www.bls.gov/news.release/pdf/famee.pdf>
- Burchinal, M. R., Howes, C. Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher-child interactions and instruction. *Applied Developmental Science, 12*, 140-153.
- Burchinal, M., Nelson, L., Carlson, M., & Brooks-Gunn, J. (2008). Neighborhood characteristics, and child care type and quality. *Early Education and Development, 19*, 702-725.
- 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.
- Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science, 6*(1), 42-57.

- Cannon, J. S., Kilburn, M. R., Karoly, L. A., Mattox, T., Muchow, A. N., & Buenaventura, M. (2017). *Investing early: Taking stock of outcomes and economic returns from early childhood program*. Santa Monica, CA: RAND.
- Card, D., & Krueger, A. (1996). School resources and student outcomes: An overview of the literature and new evidence from North and South Carolina. *Journal of Economic Perspectives* 10(4), 31-50.
- Chaudry, A., Morrissey, T., Weiland, C., & Yoshikawa, H. (2017). *Cradle to kindergarten: A new plan to combat inequality*. New York: Russell Sage Foundation.
- Clements, D. H., & Sarama, J. (2016). Math, science, and technology in the early grades. *Future of Children*, 26(2), 75-94.
- Clements, D. H., Sarama, J., & Germeroth, C. (2016). Learning executive function and early mathematics: Directions of causal relations. *Early Childhood Research Quarterly*, 36(3), 79-90.
- Currie, J. (2005). Health disparities and gaps in school readiness. *Future of Children*, 15(1), 117-138.
- Currie, J., & Thomas, D. (1995). Does Head Start make a difference? *The American Economic Review*, 85(3), 341-364.
- Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *Future of Children*, 15(1), 35-54.
- Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. *Journal of Economic Perspectives*, 27(2), 109-132.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43, 1428-1446.
- Duggan, M., Kearney, M. S., & Rennane S. (2015). The Supplemental Security Income program. In R. A. Moffitt (Ed.), *Economics of Means-Tested Transfer Programs in the United States, Volume II* (pp. 235-297). Chicago: University of Chicago Press.
- Elango, S., Garcia, J. L., & Heckman, J. J., & Hojman, A. (2015). Early childhood education. Working paper, National Bureau of Economic Research, Cambridge, MA. <https://www.nber.org/papers/w21766>
- Fisher, P. A., Frenkel, T. I., Noll, L. K., Berry, M., & Yockelson, M. (2016). Promoting healthy child development via a two-generation translational neuroscience framework: The Filming Interactions to Nurture Development video coaching program. *Child Development Perspectives*, 10(4), 251-256.
- Friedman-Krauss, A. H., Barnett, W. S., Weisenfeld, G. G., Kasmin, R., DiCrecchio, N., & Horowitz, M. (2017). *The state of preschool 2017: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.
- Fryer, R. G., & Levitt, S. D. (2013). Testing for racial differences in the mental ability of young children. *American Economic Review*, 103(2), 981-1005.
- Garces, E., Thomas, D., & Currie, J. (2002). Longer-term effects of Head Start. *American Economic Review*, 92(4), 999-1012.
- Georges, A., Malone, L., & Brooks-Gunn, J. (2012). Link between young children's behavior and achievement: The role of social class and classroom composition. *American Behavioral Scientist*, 56, 961-990.
- Gormley, W. T., & Gayer, T. (2005). Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-K program. *Journal of Human Resources*, 40(3), 533-558.
- Han, W.-J., Hetzner, N. P., & Brooks-Gunn, J. (2019). Employment and parenting. In M. Bornstein (Ed.), *The Handbook of Parenting* (3rd ed.). New York: Routledge.

## CHAPTER 2 WHAT ARE REASONABLE EXPECTATIONS FOR ECE PROGRAM EFFECTIVENESS?

- Hebbeler, K., & Spiker, D. (2016). Supporting young children with disabilities. *Future of Children, 26*(2), 185-205.
- Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*(5782), 1900-1902.
- Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the High/Scope Perry Preschool Program. *Journal of Public Economics, 94*(1-2), 114-128.
- Hill, C. J., Gormley, W. T., & Adelstein, S. (2015). Do the short-term effects of a high-quality preschool program persist? *Early Childhood Research Quarterly, 32*, 60-79.
- Hill, J., Waldfogel, J., & Brooks-Gunn, J. (2002). Differential effects of high-quality child care. *Journal of Policy Analysis and Management, 21*, 601-627.
- Hill, J., Brooks-Gunn, J., & Waldfogel, J. (2003). Sustained effects of high participation in an early intervention for low-birth-weight premature infants. *Developmental Psychology, 39*, 730-744.
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26*, 55-88.
- Hoff, E. (2012). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology, 49*(1), 4-14.
- Johnson, A., Martin, A., & Brooks-Gunn, J. (2013). Child-care subsidies and school readiness in kindergarten. *Child Development, 84*, 1806-1822.
- Jones, J. (2014, September 8). In U.S., 70% favor federal funds to expand Pre-K education. *Gallup*, Retrieved from <https://news.gallup.com/poll/175646/favor-federal-funds-expand-pre-education.aspx>
- Kalil, A., Ziol-Guest, K. M., Ryan, R. M., & Markowitz, A. J. (2016). Changes in income-based gaps in parent activities with young children from 1988 to 2012. *AERA Open, 2*(3), 1-17.
- Karoly, L. A. (2016). The economic returns to early childhood education. *Future of Children, 26*(2), 37-55.
- Kilburn, M. R., & Karoly, L. A. (2008). *The economics of early childhood policy: What the dismal science has to say about investing in children*. Santa Monica, CA: RAND.
- Klebanov, P. K., Brooks-Gunn, J., McCarton, C., & McCormick, M. C. (1998). The contribution of neighborhood and family income to developmental test scores over the first three years of life. *Child Development, 69*, 1420-1436.
- Lee, V., Brooks-Gunn, J., & Schnur, E. (1988). Does Head Start work? A 1-year follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Developmental Psychology, 24*, 210-222.
- Lee, R., Brooks-Gunn, J., Han, W.-J., Waldfogel, J., & Zhai, F. (2014). Is participation in Head Start associated with less maternal spanking for boys and girls? *Children and Youth Services Review, 46*, 55-63.
- Lee, V., Brooks-Gunn, J., Schnur, E., & Liaw, T. (1990). Are Head Start effects sustained? A longitudinal follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Child Development, 61*, 495-507.
- Lee, R., Han, W.-J., Waldfogel, J., & Brooks-Gunn, J. (2018). Preschool attendance and school readiness for children of immigrant mothers in the United States. *Journal of Early Childhood Research, 16*(2), 190-209.
- Lee, R., Zhai, F., Han, W.-J., Waldfogel, J., & Brooks-Gunn, J. (2014). Head Start participation and school readiness: Evidence from the Early Childhood Longitudinal Study-Birth Cohort. *Developmental Psychology, 50*, 202-215.

- Lee, R., Zhai, F., Han, W., Brooks-Gunn, J., & Waldfogel, J. (2013). Head Start and children's nutrition, weight, and health care receipt. *Early Childhood Research Quarterly*, 28, 723-733.
- Love, J. M., Chazan-Cohen, R., Raikes, H., Brooks-Gunn, J. (2013) What makes a difference? Early Head Start Evaluation findings in a developmental context. *Monographs of the Society for Research in Child Development*, 78.
- Ludwig, J., & Miller, D. L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *The Quarterly Journal of Economics*, 122(1), 159-208.
- Magnuson, K., & Schindler, H. (2016). Parent programs in pre-K through third grade. *Future of Children*, 26(2), 207-221.
- Marietta, G. (2010). *Lessons for PreK-3rd from Montgomery County Public Schools*. New York, NY: Foundation for Child Development.
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., & ... Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79, 732-749.
- Matthews, H., Schulman, K., Vogtman, J., Johnson-Staub, C., & Blank, H. (2015). *Implementing the Child Care and Development Block Grant reauthorization: A guide for states*. Washington, D.C.: The Center for Law and Social Policy and The National Women's Law Center.
- Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *Future of Children*, 5(2), 113-127.
- Moiduddin, E., Aikens, N., Tarullo, L., West, J., & Xue, Y. (2012). *Child outcomes and classroom quality in FACES 2009*. OPRE Report 2012-37a. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Noble, K. G., Houston, S. M., Brito, N. H., Bartsch, J., Kan, E., Kuperman, J. M., ... & Sowell, E. R. (2015). Family income, parental education, and brain structure in children and adolescents. *Nature Neuroscience*, 18, 773-778.
- Phillips, D., Gormley, W., & Anderson, S. (2016). The effects of Tulsa's CAP Head Start Program on middle-school academic outcomes and progress. *Developmental Psychology*, 52(8), 1247-1261.
- Pianta, R. C., Barnett, W. S., Burchinal, M., & Thornburg, K. R. (2009). The effects of preschool education: What we know, how public policy is or is not aligned with the evidence base and what we need to know. *Psychological Science*, 10(2), 49-88.
- Pianta, R., Downer, J., & Hamre, B. (2016). Quality in early education classrooms: Definitions, gaps, and systems. *Future of Children*, 26(2), 119-137.
- Rathbun, A., Zhang, A., & Snyder, T. D. (2016). *Primary early care and education arrangements and achievement at kindergarten entry*. (NCES 2016-070). Washington, D.C.: National Center for Education Statistics.
- Raver, C. C., & Blair, C. (2016). Neuroscientific insights: Attention, working memory, and inhibitory control. *Future of Children*, 26(2), 95-118.
- Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Smallwood, K., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly*, 63(3), 253-255.
- Reardon, S. (2011). The widening academic achievement gap between the rich and poor: New evidence and possible explanations. In R. Murnane & G. Duncan, (Eds.), *Whither opportunity? Rising inequality and the uncertain life chances of low-income children* (pp. 91-116). New York: Russell Sage Foundation Press.
- Reardon, S., & Portilla, X. (2016). Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry. *AERA Open*, 2(3), 1-18.

## CHAPTER 2 WHAT ARE REASONABLE EXPECTATIONS FOR ECE PROGRAM EFFECTIVENESS?

- Reid, J. L., Melvin, S. A., Kagan, S. L., & Brooks-Gunn, J. (2019). Building a unified system for universal Pre-K: The case of New York City. *Children and Youth Services Review, 100*, 191-205.
- Reynolds, A. J., Magnuson, K. A., & Ou, S. (2010). Preschool-to-third grade programs and practices: A review of research. *Children and Youth Services Review, 32*(8), 1121-1131.
- Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. *Journal of the American Medical Association, 285*(18), 2339-2346.
- Rock, D. A., & Stenner, A. J. (2005). Assessment issues in the testing of children at school entry. *Future of Children, 15*(1), 15-34.
- Shonkoff, J. P., & Fisher, P. A. (2013). Rethinking evidence-based practice and two-generation programs to create the future of early childhood policy. *Developmental Psychopathology, 25*(4), 1635-1653.
- Snow, C., & Matthews, T. (2016). Reading and language in the early grades. *Future of Children, 26*(2), 57-74. ?
- Steuerle, E., & Jackson, L. M. (Eds.). (2016). *Advancing the power of economic evidence to inform investments in children, youth, and families*. Washington, D.C.: The National Academies Press.
- Stipek, D., Franke, M., Clements, D., Farran, D., & Coburn, C. (2017). PK-3: What does it mean for instruction? *Social Policy Report, 30*(2), 1-22.
- U.S. Department of Health and Human Services. (2010). *Head Start impact study: Final report, January 2010*. Washington, DC: Administration for Children and Families, Office of Planning, Research and Evaluation.
- Votruba-Drzal, E. (2003). Income changes and cognitive stimulation in young children's home learning environments. *Journal of Marriage and Family, 65*(2), 341-355.
- 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.
- Yazejian, N., Bryant, D. M., Hans, S., Horm, D., St. Clair, L., File, N., & Burchinal, M. (2017). Child and parenting outcomes after 1 year of Educare. *Child Development, 88*(5), 23-39.
- Yoshikawa, H., Weiland, C., & Brooks-Gunn, J. (2016). When does preschool matter? *Future of Children, 26*(2), 21-36.
- Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M., Espinosa, L., Gormley, W., Ludwig, J.O., Magnuson, K.A., Phillips, D.A., & Zaslow, M.J. (2013). *Investing in our future: The evidence base for early childhood education*. New York: Foundation for Child Development and Ann Arbor, MI: Society for Research in Child Development.
- Zhai, F., Brooks-Gunn, J., & Waldfogel, J. (2011). Head Start and urban children's school readiness: A birth cohort study in 18 cities. *Developmental Psychology, 47*, 134-152.
- Zhai, F., Brooks-Gunn, J., & Waldfogel, J. (2014). Head Start's impact is contingent on alternative type of care in comparison group. *Developmental Psychology, 50*(12), 2572-2586.
- Zhai, F., Waldfogel, J., & Brooks-Gunn, J. (2013). Head Start, prekindergarten, and academic school readiness: A comparison among regions in the United States. *Journal of Social Service Research, 39*, 345-364.

**SECTION 1, CHAPTER 3**

**USING A SOCIAL DETERMINANTS  
OF EARLY LEARNING  
FRAMEWORK TO ELIMINATE  
EDUCATIONAL DISPARITIES  
AND OPPORTUNITY GAPS**

---

Theoma U. Iruka, Ph.D., HighScope Educational Research Foundation



The achievement gap is one of the greatest social problems in the U.S.. As currently constructed, the achievement gap indicates that White children and children from higher-income households perform better than Black, Hispanic, and Native American Indian children—and children from low-income households—on various indicators, such as reading, math, and science skills, as well as on adult outcomes later in life (e.g., health, income, educational attainment). Although most of the data substantiating this gap are gathered when children are in third grade or around age eight, there is evidence that the gap starts prior to age three. For instance, by kindergarten entry, many children from low-income and minoritized families<sup>1</sup> (e.g., Black/African American, Latino/a, non-English speaking) are months if not years behind children from White and higher-income families. We need to question how and why the achievement gap persists regardless of the academic outcomes being examined or whether we're looking at national, state, or local data. In fact, we need to stop discussing the existence of the achievement gap, or as Humphries and Iruka (2017) put it, “stop-gap gazing,” and examine the root causes of educational disparities and study how early care and education can potentially disrupt these trends.

McKinsey & Company (2009) found that not closing the achievement gap between 1983 and 1998 cost the U.S. between \$1.3 trillion to \$2.3 trillion in economic output, representing 9 to 16% of GDP. With this economic and social cost of underutilized human potential and capability, the achievement gap, which is a symptom of systemic discriminatory policies and laws, needs to be treated as a public-health crisis. In this chapter, we adapt a framework used by the public-health sector—Structural Determinants of Health—to address inequities and support the well-being of U.S. citizens at a population level (e.g., infant mortality and morbidity, teen pregnancy, or smoking) to show how early learning can address the inequities in education. To effectively eradicate disparities and inequities in early learning, we must stop gap-gazing and instead examine how systems continue to perpetuate racism and inequities that reverberate throughout the early learning system and beyond. This means examining how certain policies and laws may reduce opportunities for certain groups to thrive and meet their potential.

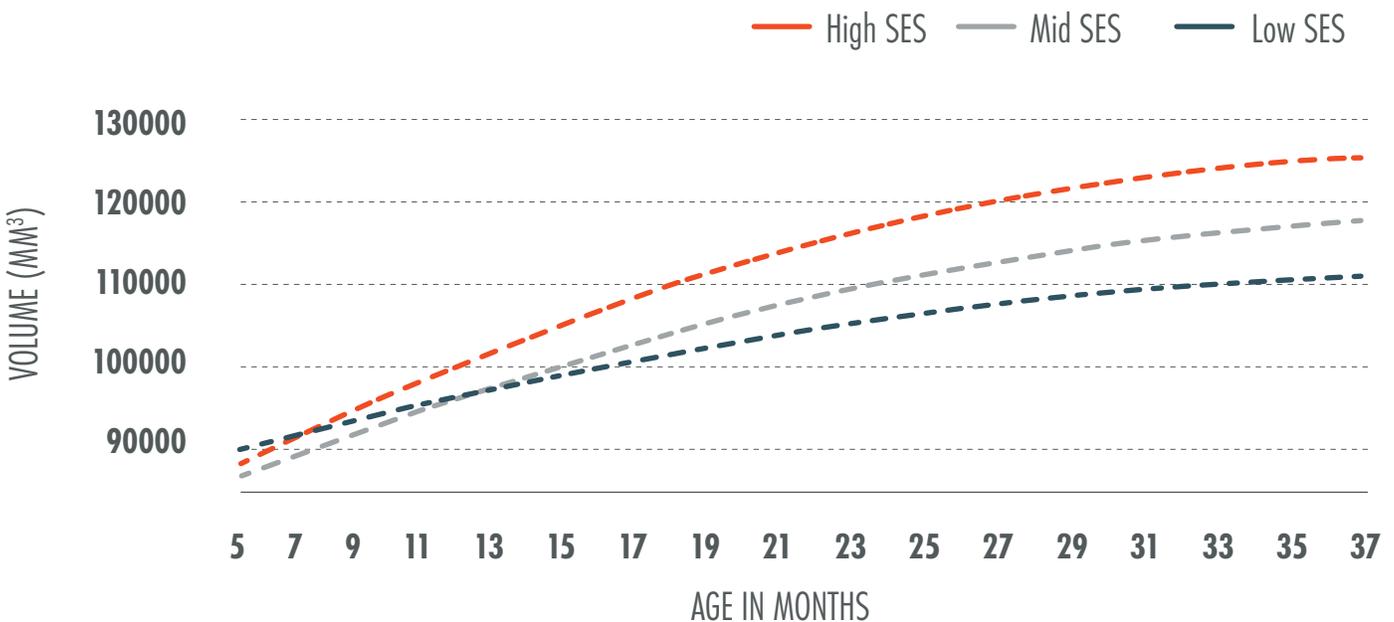
**With this economic and social cost of underutilized human potential and capability, the achievement gap, which is a symptom of systemic discriminatory policies and laws, needs to be treated as a public-health crisis.**

<sup>1</sup> Smith (2016) states that “groups that are different in race, religious creed, nation of origin, sexuality, and gender and as a result of social constructs have less power or representation compared to other members or groups in society should be considered minoritized.” People who are minoritized endure mistreatment and face prejudices that are forced upon them because of situations outside of their control. <https://www.theodysseyonline.com/minority-vs-minoritize>

## WOULD STARTING EARLY ERADICATE THE ACHIEVEMENT GAP?

What is it about the U.S. that maintains these differences and gaps? Scholars suggest that the lack of opportunities for many children of color and children from low-income households—the opportunity gap—may lead to the achievement gap. Such opportunities include access to high-quality early care and education, living in economically stable households and communities, and having enriching home- and classroom-learning environments. For example, there is a movement to ensure that children receive high-quality early learning experiences before starting school via preschool and pre-k programs, as well as programs starting at birth, such as Early Head Start and home visiting. The rationale for such programs stems from evidence showing that prior to and after birth, experience starts shaping children’s genetic potential and lays an increasingly complex foundation for learning and development. Studies show an association between poverty and cognitive development, including brain development and functioning (Hanson et al., 2013; Luby et al., 2013). Luby and colleagues (2013) find that poverty is associated with less white and cortical gray matter and smaller hippocampal and amygdala volumes (see Figure 1), which are areas that support memory, cognition, and learning. Studies examining the link between poverty and brain development, including cognitive development and executive function (EF), emphasize the dire impact of poverty, and other associated factors, such as low maternal education, single parenthood, stressful home and community environments, and poor nutrition and health, among other factors (Atkinson et al., 2015; Jeon, Buettner, & Hur, 2014). More studies should examine the impact of racism on children’s brains and health in the early years.

Figure 1. Volume of parietal gray matter in the brain across socioeconomic status (SES) groups.



Early intervention studies suggest that for children who are living in concentrated disadvantage with limited learning opportunities, experiencing enriching, high-quality early education programs at an early age may serve as a buffer that lasts a lifetime, though this is not guaranteed. When placed in a larger context, many of the children (most of them African American) who participated in the Carolina Abecedarian Project and HighScope Perry Preschool Project did not perform at the same level as White children or children from higher-income households. For example, almost a third of children from the treatment groups were arrested multiple times and did not graduate from high school, and almost two-thirds required public assistance as adults. Thus, even when looking at the best seminal early childhood programs, it seems that more than high-quality early education is needed to disrupt the influences that lead to the achievement gap and other disparities in school and life outcomes.

Although we need to strengthen the impact of early learning with other supports and structures, children who experience intellectually stimulating and enriching environments are likely to benefit from these high-quality early learning experiences—especially children from low-income households (Camilli, Vargas, Ryan, & Barnett, 2010). This is particularly critical because we know that children’s acquisition of school skills and knowledge depends on development and learning that occur long before formal schooling (Cunha & Heckman, 2008). Early school outcomes affect every area of life, including later school outcomes, family formation, child-rearing capacity, career and work preparation and stability, physical and mental health, and becoming a civically engaged, contributing member of the community and citizenry. Though access to early learning opportunities has increased, academic and social gaps by income and race/ethnicity have not been eliminated. Education scholars see some reduction in these gaps, but “at the rates that the gaps declined in the last 12 years, it will take another 60 to 110 years for them to be completely eliminated” (Reardon & Portilla, 2016, p. 12). Thus, early learning in isolation will not close the achievement gap in a timely way. Researchers, in partnership with practitioners and policymakers, must uncover and address the root causes of racial and economic disparities, and find research-based specific practices and policies that can eradicate these gaps and inequities.

## **MINORITIZED CHILDREN’S EARLY LEARNING EXPERIENCES**

---

By 2050, it is estimated, children of color will make up the majority of children in the U.S.; in 2014, children of color already made up the majority in public schools. With minority children becoming the majority, we urgently need to attend to the causes of educational disparities as early as possible. Although high-quality early learning is viewed as one strategy to ensure that children are prepared for school and life, research consistently finds that due to many stratification factors, minoritized children are at higher risk for poor outcomes than White, English-speaking children, and children from higher-income and more educated households. Race, ethnicity, and socioeconomic status (SES) are often confounded in U.S. society. Minoritized children are likely to live in concentrated poverty and disadvantage (Aud, Fox, & KewalRamani, 2010). Specifically, 34% of African American and 28% of Latino/a children and adolescents lived in poverty in 2016, compared to the 12% rate for non-Latino/a, White, and Asian children and adolescents (Koball & Jiang, 2018). African American and Latino/a youth are also more likely than

White and Asian youths to attend high-poverty, segregated schools (Urban Institute, 2019). Data from the National Center for Education Statistics consistently finds that ethnic minority children, especially African American and Latino/a children, are likely to be from single, female-headed households, live in poverty, have less-educated mothers, attend high poverty schools, and have less-educated teachers (Aud et al., 2010; McLoyd & Wilson, 1990). Additionally, low-income, ethnic-minority, and immigrant families are likely to live in racially segregated enclaves that may limit their ability to access quality early-education programs that meet their preferences (Meyers & Jordan, 2006). These disparities in social and familial characteristics are also more pronounced for dual language learners, primarily Hispanic children, compared to English speakers (Hernandez, Denton, & Macartney, 2008). Concentrated disadvantage places children at considerable risk for being less school-ready as indicated by proficiency in letter recognition and numbers and shapes, as well as for school failure and dropout (DeNavas-Walt, Proctor, & Lee, 2006; McFarland et al., 2018). If minoritized children need early learning opportunities, we must ensure that they experience the highest quality that meets their individual needs, lived experiences, and contexts.

To address the many risk factors facing disadvantaged children, federal and state programs like (Early) Head Start, Smart Start, and pre-k were developed or expanded to ensure that children placed at risk of poor school readiness and academic achievement have enriching early-childhood education programs prior to school entry (Barnett, Hustedt, Friedman, Boyd, & Ainsworth, 2007). Several studies point out that these early education programs are important for children's development and predict positive outcomes more strongly for disadvantaged children than for advantaged children (U.S. Department of Health and Human Services, 2010; Vandell et al., 2010). Not all studies have found this, however (e.g., Pungello et al., 2010), possibly because of the level of quality that disadvantaged children experience. Barnett and colleagues (2013), in a national study from the U.S. Department of Education, found that most children were in low- to moderate-quality care, but minoritized children were more likely to be in lower-quality care than were their White peers. This is concerning as many states and localities move toward universal pre-K or quality rating and improvement systems that align standards and resources for all early childhood education, including community child care (i.e., center- and home-based programs), Head Start, and pre-k programs. Although early-learning systems are being instituted, children of color and/or children from low-income households do not necessarily experience the highest quality, similar to what we see in K-12 education.

Rigorously designed early-childhood studies, such as the HighScope Perry Preschool Project and the Carolina Abecedarian Project, as well as state and municipal pre-K programs like Boston Public Schools Universal Pre-K, the North Carolina Prekindergarten Program (NC pre-K program), New Jersey's Abbott Program, and Tulsa, Oklahoma's pre-k program, have consistently and systematically shown sustained outcomes over time. But no current studies show a significant reduction in economic and racial academic disparities. For example, NC's pre-K program is a state-funded educational program for eligible four-year-olds, designed to enhance their school-readiness skills. The program operates on a school day and school calendar basis for 6.5 hours per day and 180 days per year. Local sites are expected to meet a variety of standards around curriculum, screening and assessment, training and

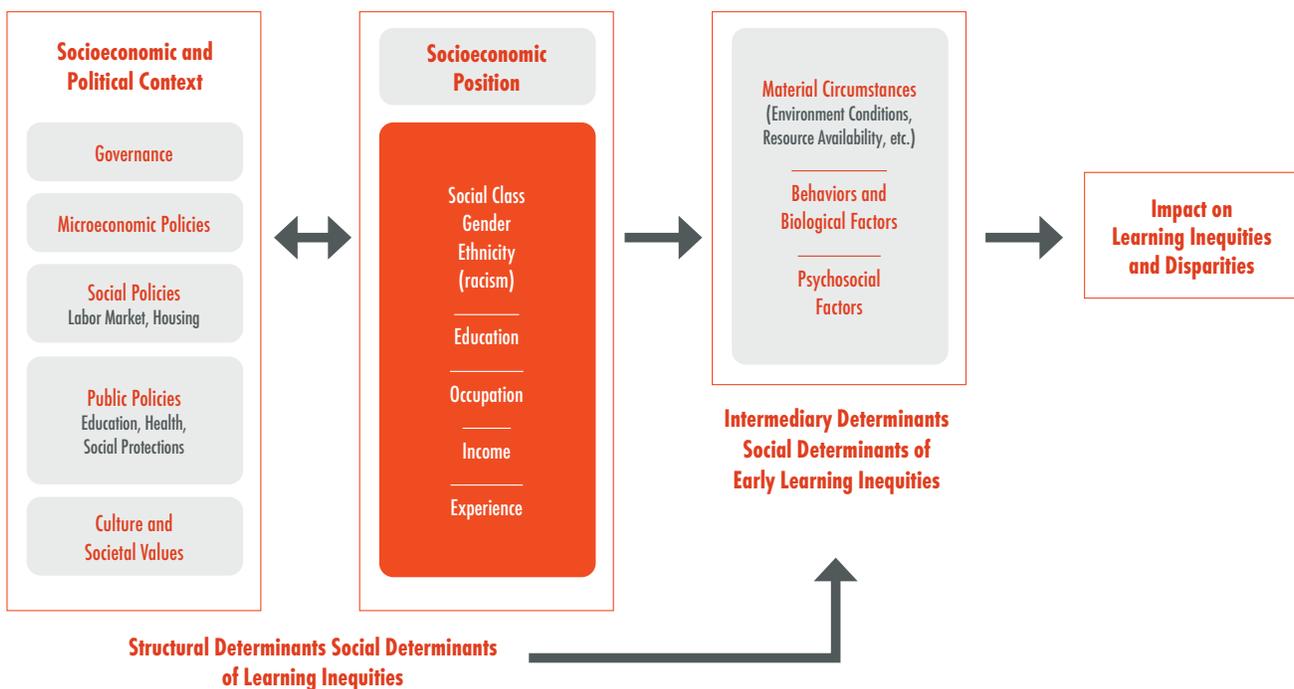
education levels for teachers and administrators, class size, adult-child ratios, North Carolina child care licensing levels, and provision of other program services. No treatment effects have been observed for language measures or teacher ratings of behavior skills at the end of kindergarten. But there are treatment effects in math and EF at the end of kindergarten for most measures, with children in NC pre-K scoring higher than matched children who aren't in the program. These effects are in the small range. Thus, while well-implemented studies show that children who get high-quality early learning do better than similar children, they don't show disruption of the achievement gap.

## SOCIAL DETERMINANTS OF EARLY LEARNING

For early childhood education to truly address early-learning disparities at the systems level, we propose adapting the Social Determinants of Health framework (SDoH) to early learning, calling it Social Determinants of Early Learning (SDoEL) (see Figure 2).

The Centers for Disease Control and Prevention defines social determinants of health as “the complex, integrated, and overlapping social structures and economic systems that are responsible for most health inequities. These social structures and economic systems include the social environment, physical environment, health services, and structural and societal factors. Social determinants of health are shaped by the distribution of money, power, and resources throughout local communities.”

Figure 2. Social Determinants of Early Learning.



Source: Centers for Disease Control and Prevention

As Figure 2 shows, the concept behind SDoEL is that socioeconomic and political contexts (e.g., social policies about housing and education) lead to individuals' socioeconomic position (e.g., education, income, or occupation), which then impacts their resources and living conditions, greatly reducing some children's opportunities to thrive. This framework is further expanded below.

### ► **Structural determinants of SDoEL**

The first structural determinant of early education is socioeconomic and political context, which include macroeconomics, public policies, and societal values. That is, the political context at the federal and state levels impacts early learning. In their *Kids Share* report, Edelstein, Hahn, Isaacs, Steele, and Steuerle (2016) find that over the past 50 years, child-focused spending grew from 0.6% of GDP in 1960 to 2% in 2015, compared to 2% to 9% for adult-focused spending during this same time period. The majority of spending on children is for Medicaid, followed by three tax provisions: the Earned Income Tax Credit, the Child Tax Credit, and the dependent exemption. Early-childhood programs, such as Head Start, are not in the top 10 for federal spending for children. Edelstein and colleagues conclude that "total federal spending on children has been fairly flat over the past four years, in real dollars. In the future, overall federal spending is projected to increase substantially, but virtually none of the additional funds will be directed toward children" (Edelstein et al., 2016, p. 11). This lack of available funding for early-childhood programs at the federal level means that fewer children, especially those most in need, may be able to access high-quality ECE programs; there may be fewer supports to ensure high-quality ECE programming; and teachers and caregivers may not be adequately compensated and supported to provide stable, high-quality, enriching early-learning opportunities. The lack of federal spending means that states and localities are spending more because they see the economic and societal value in supporting the early learning of young children. In the *State Preschool Yearbook*, Friedman-Krauss and colleagues (2018) note that although states spent more on preschool in 2017 than in 2002, going from \$2.4 billion in 2002 to over \$7.6 billion in 2017, when adjusted for inflation per-child spending during this same period decreased. This reduction in per-child spending may be due to the attempt to increase access, which rose from 14% of the four-year-old state population served in 2002 to 33% in 2017.

Beyond macroeconomics, social and public policies also have implications for early-learning disparities and equity. For example, social policies about labor have implications for early learning, such as whether being an ECE teacher and provider should be considered a career, which in turn has implications for access to adequate compensation, benefits, federal funding, etc. Currently, a wide range of early childhood advocates, practitioners, and funders are focused on creating an economically sustainable professional pathway for those who teach and care for children from birth to age 8. If successful, these efforts could ensure that all children have access to highly qualified and well-compensated ECE providers. They could also lead to increased costs to families (and possibly to programs) to provide services to children. Other policies that effect early learning opportunities include standards for programs (e.g., licensing, group size, ratio, materials, curriculum, or assessment duration), workforce (e.g., credential,

bachelor's degree, or pre- and in-service hours), and eligibility (e.g., universal or targeted). Policies about housing, the workforce, transportation, the environment, and general education, to name a few, also have implications for early-learning disparities. For example, housing policies about what constitutes adequate living conditions, standards for renters and landlords, and availability of affordable housing, etc., affect children's well-being. Policies about affordable housing or the lack thereof could bear on who can live in a particular community. This impact may be particularly pronounced for low-resourced families. Coupled with transportation policies, such policies could impact a communities' ability to ensure that residents are gainfully employed, which affects the community's tax base—a potential source for early-learning funding.

When SDoEL is overlaid with critical race theory (CRT),<sup>2</sup> we can recognize that race and racism are enduring and pervasive in the U.S. and that power structures lead to systematic inequities (Matsuda, Lawrence, Delgado, & Crenshaw, 1993). Recognizing that race permeates the fabric of the U.S. and the lived experiences of minoritized groups, and finding ways to systematically address racism in education, including early learning, is pertinent to culturally responsive and sustaining practices and pedagogy. When we examine macroeconomic policies, such as housing and environmental policies, as well as their historical ramifications, we see that Black people and other people of color are often disenfranchised and marginalized. The U.S. policies that barred Black families from owning and renting in particular areas have resulted in Blacks living in segregated enclaves that are characterized by more poverty, crime, dilapidated housing, low-resourced schooling, low-quality air and water, and limited employment options. This residential segregation has had a detrimental impact on the opportunities of Black people for generations, including those who are highly educated and middle-class (Massey, Condran, & Denton, 1987). Segregation also affects early childhood. Over 50% of Black and Hispanic preschool children in public school-based programs attend racially segregated schools (Urban Institute, 2019). Reid, Kagan, Hilton, and Potter (2015, p. 5) note that “most children in public preschool programs attend economically segregated programs that are often segregated by race/ethnicity as well.” Studies have shown that programs serving high proportions of children of color and children from low-income homes are less enriching and engage in more routine-based activities (Early et al., 2010), further exacerbating early-learning disparities. Thus, the U.S. historical and contemporary culture of limited opportunities for children and families of color has lifelong implications for families' socioeconomic position, which directly impacts children's early learning and later outcomes, and the opportunities provided to them.

Families' socioeconomic position represents another structural determinant of early education; it includes social class, gender, ethnicity, education, occupation, and income, and is likely determined by the U.S. socioeconomic and political context. Policies about labor, employment, housing, and education, etc., have a direct impact on

---

<sup>2</sup> Critical race theory came out of legal scholarship that recognizes that racism is engrained in the fabric and system of the American society, that institutional racism is pervasive in the dominant culture, and that power structures are based on white privilege and white supremacy, marginalizing people of color and others due to sex, class, national origin, and sexual orientation.

families' socioeconomic position, and this impact varies based on their social class, gender, and ethnicity. For example, policies about the need for bachelor's degrees or higher for certain positions (e.g., teaching) could lead to stratification based on race and economic background; such stratification has long-term impacts, especially when many of these positions may have livable wages and salaries, benefits, and pensions. These policies are manifested in the education field. National data indicate that although minoritized children make up the majority of public school students, White teachers make up over 72% of the preschool and kindergarten teaching workforce (Black teachers make up 18% and Hispanic teachers even less; <https://datausa.io/profile/soc/252010/#demographics>). Taking a CRT perspective, the systematic barriers for people of color to access and afford higher education, which then influences the type of positions they are qualified for, shows how inequities are maintained through policies that directly impact access and opportunities for families of color and their children. Even when people of color qualify for particular positions, they are likely to earn less than their White counterparts. In the 2018 Early Childhood Workforce Index, Whitebook, McLean, Austin, and Edwards (2018) found that black center-based teachers are more likely than teachers from all other racial/ethnic groups to earn less than \$15 per hour, which has implications for their socioeconomic position. A pay rate of \$15 per hour results in an annualized salary of \$31,200 (without benefits), keeping one's income above the U.S. poverty threshold of approximately \$24,000 for a family of four.

### ► **Intermediary determinants of SDoEL**

Based on the conceptual framework of SDoEL, structural determinants influence individuals' and families' processes. For example, families' socioeconomic positioning affects their material circumstances (e.g., food, housing, and work conditions), behavior and biological functions, and psychosocial factors (e.g., stress). Scholars have found that families' investments and stressors are possible explanatory factors linking socioeconomic status to children's school readiness (Iruka, LaForett, & Odom, 2012; Mistry, Benner, Biesanz, Clark, & Howes, 2010; Raver, Gershoff, & Aber, 2007). Specifically, the family-investment model postulates that parents with more income, time, and education are able to provide enriching learning opportunities and resources that support children's learning. But families with fewer economic resources experience a lot of stress, which increases depression and detachment, minimizing the quality of interactions and relationships with their children and having a detrimental impact on children's development and learning.

Socioeconomic positioning is associated with early-learning disparities due to other factors and conditions in the environments in which children are born, live, learn, and grow up that affect the quality of their development and the risks they face. The social, economic, and physical conditions of children's homes, communities, and early-education settings affect children's learning opportunities. For example, children's economic condition determines the early-learning environments children can access. That is, the quality of children's environments at home or outside the home often determines the quality of the learning environments (e.g., safe, nurturing, and enriching) and interactions (e.g., responsive and language-rich) they are likely to experience. Higher-income families likely can afford better-quality environments (Barnett et al., 2018). In comparison to those with limited opportunities,

children who experience high-quality learning environments in their daily lives are likely to have better opportunities that set the foundation for school readiness and a better school trajectory. Access to other resources—such as safe and affordable housing, reliable transportation, employment, safe and nonviolent communities, healthy foods, health services, and environments free of life-threatening toxins—impacts families’ and children’s mental health and functioning, which in turn affects children’s learning and development trajectory over time. This pattern aligns with Bronfenbrenner and Morris’s (2007) bioecological framework, which emphasizes that children’s development hinges on multiple contexts and systems. Indeed, research has shown the interconnection between community contexts and child outcomes (e.g., Dupéré, Leventhal, Crosnoe, & Dion, 2010). First, “collective norms and socialization, as well as the relative level of stress and support in the neighborhood, are primary ways in which neighborhood characteristics may influence parenting and, in turn, achievement” (Dupéré et al., 2010, p. 3). Second, “community socioeconomic characteristics shape the composition and quality of local institutions whose mission revolves around children’s cognitive growth, such as child care and school, and that this, in turn, influences achievement. [In essence], neighborhood financial, human, and social capital all influence the strength and vitality of neighborhood learning institutions” (Dupéré et al., 2010, pp. 4-5).

## HOW CAN THE SOCIAL DETERMINANTS OF EARLY LEARNING STRENGTHEN EARLY EDUCATION TO ADDRESS DISPARITIES?

To maximize the benefits of homes and communities and buffer children from negative factors, ECE environments, systems, and classroom environments can serve as intermediaries. That is, to reduce economic and racial disparities in the early years, ECE can serve as a place-based conduit and centralizing institution to ensure that children receive early-learning opportunities that take into account the structural determinants impacting their learning. In particular, ECE must attend to the racialized U.S. context, in which children from low-income households and minoritized children and their families face more challenges and inequities than higher-income and White children and families. Garcia Coll and colleagues (1996, p. 1895) emphasize the notion that to really deliver on the promise of early childhood to equalize opportunities for minoritized children, we must consider how environments like ECE can buffer children from the effects of low and marginalized socioeconomic positions (e.g., social class, gender, and ethnicity) that lead to segregated, inadequate communities caused by “pervasive social mechanisms of racism, prejudice, discrimination, and oppression.” Although individuals may have assets directly linked to children’s learning and development in the early years, we must acknowledge the systematic influences that set children’s trajectories based on factors outside their control (e.g., race/ethnicity, language, zip code, quality of child care, and ECE policies).

**To reduce economic and racial disparities in the early years, ECE can serve as a place-based conduit and centralizing institution to ensure that children receive early-learning opportunities that take into account the structural determinants impacting their learning.**

Thus, to truly ensure that all children have access to and equitably benefit from high-quality early-learning practices, and to address educational disparities, we need to consider these social determinants. We need to build on existing birth-to-age-five programs and systems with attention to SDoEL and how structural factors impact children's school readiness and later outcomes. Some of these programs and systems include home-visiting programs, birth-to-age-three programs (e.g., Early Head Start), and quality rating and improvement systems, which I discuss below.

### ► **Birth-to-age-three and home-visiting programs**

Evaluations of early intervention programs focused on infants and toddlers have shown mixed results, especially in regard to children's cognitive, language, and socioemotional outcomes. One example is Early Head Start (EHS), a two-generation program designed to provide high-quality child and family development services to low-income pregnant women and families with infants and toddlers. In 1996, the Early Head Start Research and Evaluation Project, involving 3,001 families at 17 sites, found some positive, albeit small effects for children's cognitive and receptive language. The program was found to have more favorable impacts on children's socio-emotional development in regard to their interactions, attention, and negativity with parents during play, as well as how aggressive their parents reported them to be. When children in EHS were examined two years later during the preschool years, evaluators still found significant impacts for socioemotional behaviors in the areas of behavior problems and approaches to learning; with the exception of a positive impact for Spanish-speaking children's receptive language, there were no other achievement-related outcomes.

The recent Home Visiting Evidence of Effectiveness study funded by the U.S. Department of Health and Human Services (<http://www.acf.hhs.gov/programs/ecd/home-visiting>) provides evidence of a positive and long-term impact from various home-visiting programs that focus on improving the quality of the home environment and increasing positive parenting. Over 30 home-visiting programs have been found to be evidence-based, as determined by at least two impact studies. The outcomes these home-visiting programs focus on included child health, child development and school readiness, family economic self-sufficiency, linkages and referrals, maternal health, positive parenting practices, reductions in child maltreatment and juvenile delinquency, family violence, and crime. Several of the programs' findings have been sustained over time and replicated with other samples, but we still need to ensure that these programs are lifting families and children out of poverty and setting them on a path to economic stability and life success (Avellar et al., 2016).

These birth-to-age-five programs produce the following evidence:

- starting sometime in the first five years of life is positive, especially for children from low-resourced households;
- home-visiting programs that start before or right after birth are beneficial for both children and parents across many outcomes;

- with the exception of small-scale, rigorously controlled early-intervention programs (e.g., Perry Preschool, Abecedarian), findings about the long-term impact of preschool/pre-K programs and the closing of the achievement gap have been limited and inconsistent; and
- many preschool programs show attenuation of findings over time, as early as the following year.

Various scholars have noted “fadeout” following these early experiences (Barnett, 2011). Some argue that fadeout is due to the minimal impact of early-childhood experiences (Whitehurst, 2013), while others suggest that it may represent a “catching up” of those who did not experience high-quality early education, or that there may be a “sleeper effect” of persistent impact evident later in life (Barnett, 2011). For example, some have argued that the impact of early-childhood programs such as Head Start may not be sustained because of the low-quality schools that Head Start children are likely to transition into (Currie & Thomas, 2000; Garces, Thomas, & Currie, 2002). Another theory is that teachers are focusing on children with the lowest skills to help them catch up, and these may be children who did not experience high-quality early learning. Fadeout indicates a need for continued alignment of educational programs beyond five years (e.g., birth-to-8 initiatives), but could also indicate that the things most predictive of school and life are not appropriately captured (e.g., persistence or social-emotional learning; Heckman & Karakapula, 2019).

### ► Prenatal to grade 3<sup>3</sup>

Research tells us that the brain develops most rapidly in the earliest years; that enriching early-learning experiences are critical for children’s long-term success (Shonkoff et al., 2012); and that longer-term benefits and outcomes both for the child and for society are seen with multi-year, high-quality programs across the early grades, at least based on small controlled studies (Vandell et al., 2010). The National Research Council Report From Neurons to Neighborhoods (Shonkoff & Phillips, 2000) makes the compelling case that the earliest years—birth through the primary grades—are critical to the long-term educational and life success of all children. And evidence suggests that if quality interventions and programming are provided, gains in cognitive and socioemotional skills may be greatest for children who are farthest behind (Reynolds, Temple, Ou, Arteaga, & White, 2011; Shonkoff et al., 2012). As I discuss above, evaluations from early intervention programs show that starting early does matter, especially with home-visiting and high-quality early education programs. With the exception of small longitudinal studies, there have been mixed findings regarding the longer-term impact of preschool programs<sup>4</sup> or birth-to-age-five programs. Thus, as a way to consolidate the impact of high-quality early experiences, especially for children placed at risk for poor

<sup>3</sup> “PreK-3rd Grade” is used interchangeably with “P-3.” Both terms are intended to reflect the importance of aligning across birth-to-five (0-5) and K-12 classrooms and systems.

<sup>4</sup> We use preschool to denote programs or services provided to children from birth to age 5. Pre-k is used to refer to programs offered to four-year-old children or a year prior to children’s entry into kindergarten.

outcomes, the field has focused on alignment between preschool and the early elementary years (Ma, Shen, Krenn, Yuan, & Hu, 2015). This has resulted in many programs, strategies, and initiatives focused on prenatal to grade 3 (P-3) to better align “education practices (teachers), education policies (principals), and education standards (curriculum, instruction, assessment, and professional development) [and] make horizontal connections within each grade level and vertical connections across different grade levels in order to create seamless logical transitions that ensure academic and social success for students” (Ma et al., p. 1069). An indication of this approach is also seen in the establishment of the National P3 Center: “[T]he vision for P-3 approaches is to improve the quality and coherence of children’s learning opportunities, from the experiences children have in early learning (including pre-K, Head Start, child care, and other early-learning opportunities before—or “pre”—formal entry into school) and extending through elementary school” (<http://depts.washington.edu/pthru3/>). The premise for P-3 is that coherent, high-quality instructional approaches across this age-and-grade span will result in positive outcomes for children throughout their early years, and an increased likelihood that children will be minimally on track by the end of third grade toward school and life success.

### ► **Quality rating and improvement systems**

The fact that most children are likely to be in community-based programs, especially home-based and informal settings, suggests a need to establish early-learning systems that systematically address the structural determinants of early-learning inequities and disparities. The desire to ensure that all early learning programs are of high quality by operating under the same standards and expectations has led to the implementation of quality rating and improvement systems (QRISs). In developing these systems, state and local policymakers have used research linking high-quality early childhood education and children’s outcomes in developing QRISs. The idea is to ensure that all children, especially disadvantaged children, are attending high-quality education programs during their early years. Nearly all state QRISs measure staff training and education and assess the classroom or learning environment. States differ on whether and to what extent they include parent-involvement activities, business practices, child-staff ratios, or national-accreditation status. QRISs serve multiple purposes, one of which is to provide a standard way to rate program quality based on multiple criteria and make that information available to parents. The assumption underlying this function of a QRIS is that parents often lack good information about program quality and that if such information were available, they would be more likely to choose higher-rated programs. As a result, lower-quality providers would have an incentive to either improve the quality of their program or to leave the market (Zellman & Perlman, 2008). Also, QRISs represent a systematic approach to providing a range of technical assistance, resources, and incentives for programs to improve quality. Such efforts include consultation around quality improvement, increased investments for professional development scholarships, microgrants for other targeted quality-improvement efforts, and in some instances higher levels of subsidy payments for more highly rated programs. The goal is to foster and support providers’ efforts to improve the quality of care they provide. Thus, QRISs attempt to improve quality by affecting both the demand for high-quality care and the supply of such care.

Of course, their success rests on their ability to accurately identify and measure key aspects of quality and on the willingness of providers to participate in a rating system (Zellman & Perlman, 2008).

The evidence that QRISs lead to better program quality, child growth, and school readiness is mixed. A recent compilation of validation studies from 10 states found the following (Tout et al., 2017):

- (a) levels of quality in the medium range;
- (b) significant, albeit small associations between ratings and observed quality in center-based programs, with differences in the areas of environments, interactions, and activities in ECE programs at different rating levels;
- (c) ratings generally distinguish between lower and higher quality, but no support for the idea that each level of a QRIS reflects a meaningful difference in quality from other levels; and
- (d) inconsistent evidence of small positive associations between QRIS measures and child outcomes, mostly for ratings of social-emotional development and assessment of executive function.

The differences in system designs across states make it difficult to draw general conclusions from the eclectic validation studies. Furthermore, most of the states have few programs in the highest level of quality, resulting in two categories of quality, low and high, that may impact the links to child outcomes. Other limitations include the focus on three- and four-year-old children compared to infants and toddlers within a small time frame of about six months; the need for other measures of children's learning and development; limitations of quality measures that may need more calibration and refinement of area; and use of classrooms to indicate center-level quality.

QRISs have the potential to be a conduit for early learning, family support, and health and well-being to ensure that children of color, children from low-income households, and children from other marginalized communities have equitable opportunities to thrive and be successful. But many QRISs are voluntary, indicating that most programs serving children with high needs may not participate unless mandated (e.g., as part of a subsidy system). Programs with the highest standards and best workforces may not participate in a system that is accessible by many families of color or low-income families. Last, the standards in QRISs have not been considered with the SDoEL framework in mind. For example, how are the standards ensuring that these systems are not privileging certain groups and penalizing others? Are programs serving children who have the greatest needs and who are in the neediest communities being provided with resources to meet their needs? To what extent is segregation being addressed to ensure that families have diverse high-quality choices to meet their needs and ensure their children are excelling?

## SETTING A RESEARCH PRACTICE-POLICY AGENDA FOR ECE PROGRAMS AND SYSTEMS TO DELIVER ON THE PROMISE TO PROVIDE EQUITABLE OPPORTUNITIES USING THE SDOEL FRAMEWORK

---

*(1) ECE research must consider racism and discrimination using the SDoEL framework.* For too long, most ECE research has indicated that many children of color and children from low-income households are not prepared for school and need early care and education programs. Unfortunately, most of the research, especially about children of color and their families, has been done with a deficit perspective, without consideration for the social determinants that lead to the disparities witnessed even after interventions. The results have often shamed and blamed children, families, and communities for low scores on language and cognitive assessments without considering the historical legacy of racism and discrimination and white supremacy that couches all aspects of early learning. Not even minimal consideration has been given to the resilience and perseverance of children of color and their families, who continue to thrive even when they are subjected to systems and institutions that limit their opportunities and don't consider their assets. When it comes to minoritized populations, are we asking and answering the right questions in the right way? Are there areas in which children of color and other marginalized groups are overperforming that are not considered or addressed (e.g., oral language and storytelling, or bilingualism and biculturalism)? For instance, one would assume that children who have to learn to operate one way at home and another way at school must have strong cognitive skills, but these skills are not captured in discrete assessments, nor is credit given for children who have a home language or dialect and then have to switch to another dialect and language in other settings. Thus, in addition to examining how ECE can help minoritized and marginalized children, research needs to examine how structures and policies promote or hinder families' and communities' ability to thrive and promote children's learning. Research can also help determine what standards can ensure that all children equitably thrive, rather as opposed to standards based solely on Eurocentric ideals of what is good and appropriate. A sole focus on what occurs in the classroom without an understanding of how macrosystems and policies impact it does not help increase ECE's impact—hence the importance of the SDoEL framework to guide research studies.

*(2) Engage in cross-sector collaboration with the SDoEL framework.* Inequities and disparities are not created because parents are “lazy” or “uncaring.” Rather, structural features work in concert to impinge on the abilities and processes of families and communities; these features include policies that increase poverty and reduce economic mobility, housing and education patterns that maintain low-income segregation, and limited transportation options that restrict the ability to find and maintain employment. Thus, while parents may be able to support their child's healthy development and learning, factors beyond their control (e.g., economic stress, community safety, environmental toxins, or unstable and non-standard unemployment) may limit this ability. As with health disparities, similar structural and process determinants

lead to early-learning disparities and inequities. The root causes of these disparities and inequities often lie in historical and contemporary policies and structures (e.g., education, housing, employment, health systems, public safety, income, and wealth), and some of them are vestiges of U.S. institutional racism. These root causes have not been focused on or studied in ECE research. Although early-learning programs and systems have been shown to mitigate some challenges in the home environment by providing children with consistent, sensitive, and cognitively enriching learning opportunities, such opportunities are not always accessible or of high quality, especially for low-income and minoritized groups, and especially for Black children. Thus, we need to examine how supports can be effective for children and their families, for example, by understanding how health systems and family systems interact with ECE systems to promote positive and optimal child development and learning.

*Potential steps for engaging in cross-sector collaboration:*

- build a coalition with multiple agencies and organizations that intersect with the SDoEL (e.g., family support, early learning, education, housing, workforce, child welfare, and criminal justice)
- identify coalition leaders and potential ways to integrate work into current funding or organization infrastructure
- determine collective impact outputs (e.g., healthy and safe early childhood, kindergarten readiness, third-grade reading, family stability, diverse schools, livable wages, and affordable housing)
- develop a data process and system to monitor challenges and changes
- develop a continuous quality improvement process at multiple levels
- develop policy changes aligned with communication strategy and resource needs

**(3) Using the SDoEL framework for ECE systems and workforce.** The bulk of this chapter focuses on the social determinants experienced by families. But we need to recognize that the ECE workforce is also impacted by the same systems that lead to early-learning disparities. Studies have shown that many ECE professionals, particularly those working in community-based programs, are living at or below the poverty level and seek social benefits and services similar to those sought by the families they serve. Thus, they are likely experiencing economic stress and poverty, which affect the quality of their interactions with children and the instruction they provide in the classroom, as well as turnover (i.e., instability), which has also been associated with quality. Poverty and stress are more likely to impact ECE professionals who are members of historically marginalized groups and, by extension, children of color and those from low-income households. Furthermore, these programs and providers may have less access to resources. Rather than focusing solely on the challenges experienced by children in programs and schools, we also need to pay attention to the challenges experienced by ECE professionals.

This means that ECE programs and systems may need to examine the demographic makeup not only of children and families, but also of educators and leaders. It may also mean advocating for more resources for programs, as well as economic resources for ECE professionals, to ensure that social determinants are not being perpetuated throughout the system. For example, the Early Childhood Workforce Index (Whitebook et al., 2018) indicates that teacher assistants and teacher aides closely mirror the children they serve in race, ethnicity, and language, in comparison to lead teachers and education leaders. These lower-level positions, while important for children’s experiences, also maintain a status quo that preserves inequities in families and communities of color. Thus, we should pay attention to leadership opportunities in ECE programs, schools, and systems, for many reasons. One is the need to have diversity of minds and experiences to strengthen programs and schools, and to create a different narrative about the value of people of color; another is to ensure that upward mobility is equitably available.

***(4) Integrating CRT and culturally responsive pedagogy (CRP) in early-learning systems and programs.***

Because economic and racial disparities are part of the social and educational challenges of our lifetime, we need to understand how early-learning systems and programs could help alleviate some of the root causes that maintain inequities. Because the lives and learning styles of children of color are often marginalized, early learning program leaders and educators could fruitfully examine the extent to which programs, schools, and systems can better incorporate CRT and CRP in their standards, assessments, curricula, learning-environment structures, policies, accountability systems, quality indicators, etc. It is critical that early-education systems, programs, and educators eliminate racism and inequities in structures and processes. Important questions include: Whose standards are we using, and what is the evidence and relevance for underserved and marginalized children? For example, does emotional support look the same across different communities? How does bias look in observational assessments?

Early learning is viewed as a potential strategy to mitigate gaps by income and race/ethnicity. But at the rate we are going, it would take about 100 years to eliminate the achievement gap, and even that is not guaranteed. Racism, discrimination, and inequities are complex issues. As more children are living in low-income homes, especially among minoritized populations, the challenges of living in low-resourced and historically segregated communities affect children’s early learning and eventually their later development. With minoritized children becoming the majority, early-learning programs and systems need to consider whether and how ECE programs and systems are integrating a culturally responsive perspective that rejects bias. This perspective is particularly important when studies continue to show that links between classroom quality and child outcomes are minimal—possibly because we have paid too little attention to how individual children—especially underserved and marginalized children whose culture and lived experiences are often not considered in ECE programs’ and schools’ curriculum—are experiencing the learning environment. For example, how is the lived experience of a Black boy in the rural South considered in the

implementation of curricula, activities, literacy tools, interactions, and assessments? Or is it assumed that all children just require the same amount and type of sensitive and cognitively enriching interactions and instructions, without acknowledging their health, family and home condition, community environment, or narrative about their race or neighborhood? Even more important, what roles do racism and discrimination play in the lives and early-learning experiences of minoritized children, and their later outcomes? Understanding and clarifying the empirical links between racism and discrimination could set the course toward ensuring that programming and practices consider these issues in all aspects, in the same way that trauma-informed care addresses toxic stress. One can't address what one does not fully acknowledge.

With this perspective in mind, Brown-Jeffy and Cooper (2011) propose a culturally relevant pedagogy that ECE professionals should consider in all aspects of their work. The model comprises five themes: (1) Identity and Achievement, (2) Student Teacher Relationships, (3) Equity and Excellence, (4) Developmental Appropriateness, and (5) Teaching the Whole Child. It requires teachers to understand cultural differences between them and their students, as well as their own potential biases and stereotypes about their students.

In the Identity and Achievement area, the authors stress the notion that everyone has a multicultural identity; however, race plays a central role in many people's identities. Thus, we have to recognize the stereotype or bias about individuals from an ethnic minority group and how that may impact the quality of instruction and interaction; we also need to recognize the importance and value of affirming different cultures and lived experiences.

The Student Teacher Relationship is the mechanism that supports children's active engagement in a classroom or program, especially when children spend many hours per day over months and years with the same teacher or teachers. These relationships create a classroom culture that extends into children's lives, shaping how they view and interpret the world, others, and themselves. Equity and Excellence focuses on the notion that teachers (and systems) have to provide what children and families need in multiple forms, rather than focusing solely on equality.

Developmental Appropriateness emphasizes children's learning zones (what they have mastered and are on the verge of mastering) and considers the assets children bring as well as an understanding of how the remnants of racism may impact and influence children's learning and development (e.g., viewing children's home language, such as African American English vernacular, as evidence of unintelligence). Teaching styles should be integrated with children's learning styles, and teachers should be aware that some children's learning styles may not be viewed favorably from a noninclusive white, Eurocentric perspective.

Last, *Teaching the Whole Child* emphasizes the importance of recognizing that culture in all aspects of children’s systems—from the home to the community to society—causes them to receive, respond, perceive, and prioritize meaning and behavior in different ways. In essence, “teaching the whole child will require not only that teachers recognize, understand, and intentionally acknowledge cultural group behaviors, but also observe and interact with students as individuals” (p. 77). With this framework in mind, research can help ascertain the extent to which these five themes enhance minoritized children’s educational experience.

*(5) Implementation should consider the quality of inputs and structures.* Due to their various root causes, early-learning disparities are complex. They require a complementary, cohesive system and approach that asks the right questions, conducts the right research, and implements the evidence in a cohesive way. At present, advantaged families can access programs and schools that provide high-quality, personalized instruction with highly educated, stable, and cognitively stimulating educators. These families can also create separate learning systems, schools, and programs that maintain privilege and the status quo. For example, Montessori and Reggio-inspired programs are often found in highly-affluent communities, though these pedagogical approaches were created for children from low-income and challenged families. In these programs, teachers are expected to be fully credentialed, with at least a bachelor’s degree, and go through several years of preservice practicums with continued in-service work to maintain their credentials. Most teachers stay for decades, and their leaders often embrace the autonomous nature of teaching and create an affirming and comfortable work environment. A level of standards is expected regardless of state, city, or locality, and families are willing to pay the necessary amount for such an educational experience. Alternately, publicly funded programs and schools are subject to federal and local policies and funding, as well as standards that may not take into account the needs of communities and families or the available resources or capacities. Most early learning programs cannot afford the highest quality staff, or the resources needed to ensure that quality is sustained over time, especially with their relatively high turnover rates. Although we have evidence-based curricula, we have no general pedagogy about how best to teach and support young children, especially children with diverse needs, learning styles, and experiences. Early-learning standards and expectations vary across and within states, creating further challenges about what it takes to create and maintain a high-quality early learning system and program. Even the measures and systems created don’t provide precision about the actual quality of a program and how to increase it.

Rather than focusing on points and ratings—although they may be helpful for communicating with families, educators, and policymakers—we need to focus on the quality and capacity of the workforce to provide equitable learning opportunities. We need national standards about what early learning should and can be expected to provide across diverse settings and groups. We need to gauge the cost per child of providing quality early-learning experiences and ensuring that equity rather than equality is the approach taken with funding and resources. We need to encourage systems to align workforce, resources, and data to meet the needs of children’s learning and development. Implementation of high-quality early learning

should focus not only on classroom instruction, but also on the infrastructure that supports processes, including leadership, funding, standards and regulation, data, and partnerships across programs. We should pay attention to how these factors create barriers to or disincentives for equitable early-learning opportunities. For example, do licensing or standards ban blended classrooms or mixed ages, which may be beneficial for some groups of children and have implications for the types of programs that would be allowable? Should these programs be expected to prevent or reduce learning disparities prior to formal school entry in isolation?

## CONCLUSION

---

Early learning is a promising approach, but it is impacted by social determinants that maintain inequities and thus ensure disparities. These structural factors limit resources and supports that directly impact children's outcomes, especially for low-income and minoritized children and their families. The return on investment and effectiveness of early-learning programs were primarily established with Black children; however, Black children are still likely to perform more poorly on almost every marker of learning and optimal development than their White peers. Furthermore, they are likely to experience an intractable cycle of racism and discrimination that has not been fully fleshed out and examined in ECE research. To truly ameliorate early learning inequities and disparities, we must recognize systems that invisibly maintain and perpetuate inequities from housing to education; build cross-sector collaboration and partnership through a racial equity–research lens; and develop a collective birth-through-elementary school (if not, arguably, birth-through-career) strategy to ensure that all children, regardless of race, ethnicity, language, gender, or community, have the opportunity to reach their potential. Early-learning programs and systems are the first formalized institutions that children and their families likely experience; thus, they should take charge in creating a culture that ensures racism and inequities are considered and addressed, in coordination with other sectors.

For ECE programs to meet these expectations, the ECE field has to engage in more thoughtful, meaningful, and racially responsive research focused on understanding the causes and solutions for learning disparities and gaps. This will require the ECE research community to take an equity perspective that includes diverse voices and perspectives—especially those from minoritized communities—to examine how social and structural determinants impact children's outcomes. Although researchers may be interested in microlevel factors, such as classrooms and families, we need a critical examination of how macrostructures and policies may impact such microlevel systems and thus children's outcomes. The “color-blind” approach to research by “controlling” for race, ethnicity, language, and gender must be minimized because it undermines experiences based on these social markers. Furthermore, scholars must undertake interdisciplinary ECE research that engages multiple sectors (e.g., education, health, social work, and workforce development) and disciplines (neurobiology, public health, urban planning, economics, medicine, and implementation science). The solution to pernicious disparities and inequities must be thoughtful, with attention to history and with collaboration from multiple disciplines. All children deserve to start off right and have an equitable opportunity to learn and thrive.

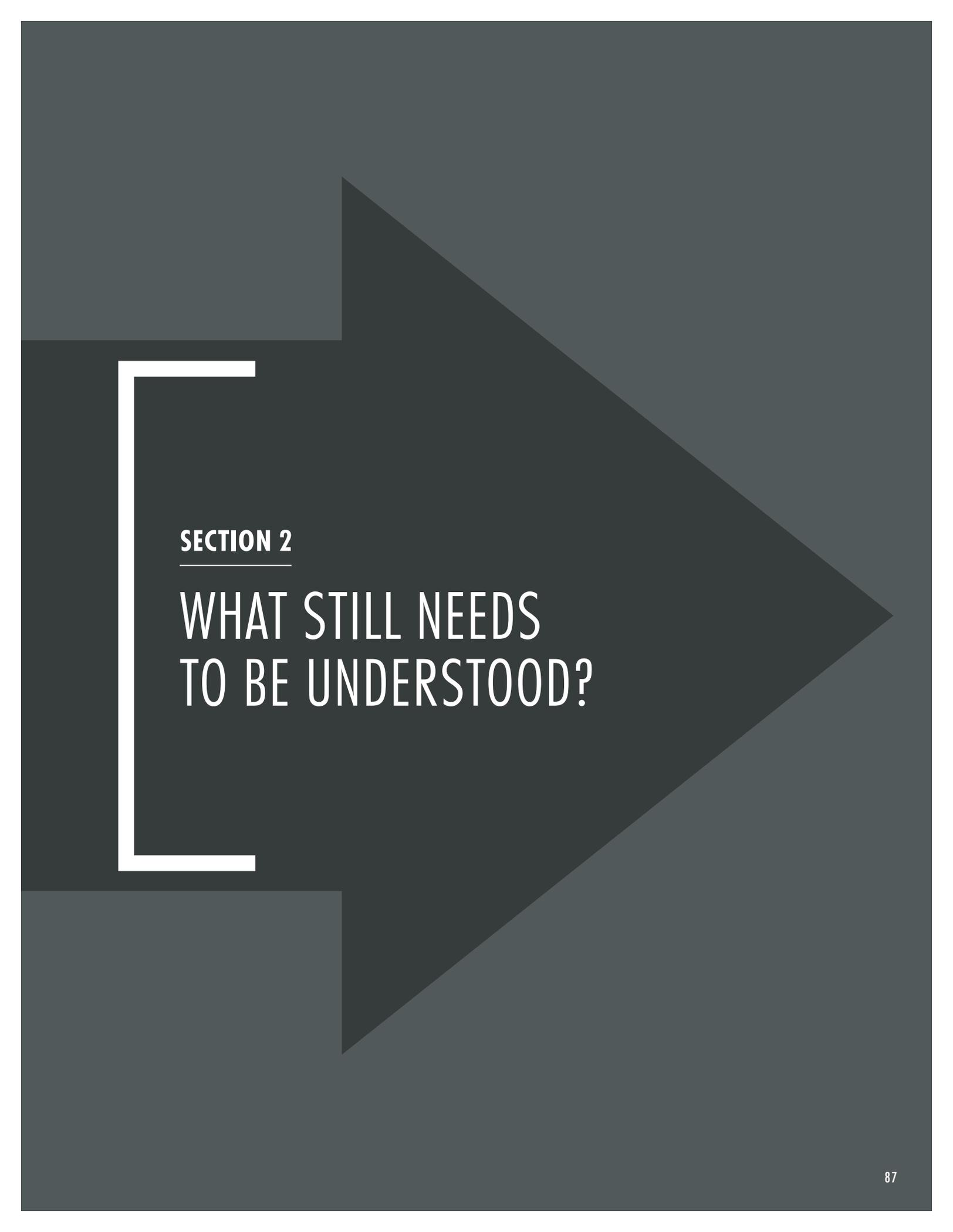
## References

- Atkinson, L., Beitchman, J., Gonzalez, A., Young, A., Wilson, B., Escobar, M., . . . Villani, V. (2015). Cumulative risk, cumulative outcome: A 20-year longitudinal study. *PLoS ONE*, 10(6), e0127650. doi: 10.1371/journal.pone.0127650
- Aud, S., Fox, M.A., & KewalRamani, A. (2010). *Status and trends in the education of racial and ethnic groups (NCES 2010-015)*. U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Avellar, S., Paulsell, D., Sama-Miller, E., Grosso, P.D., Akers, L., & Kleinman, R. (2016). *Home visiting evidence of effectiveness review: executive summary*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.
- Barnett, W.S. (2011). Effectiveness of early educational intervention. *Science*, 333(6045), 975-978. doi: 10.1126/science.1204534
- Barnett, W.S., Carolan, M., & Johns, D. (2013). *Equity and excellence: African-American children's access to quality preschool*. New Brunswick, NJ: Rutgers University, National Institute for Early Education Research, Center on Enhancing Early Learning Outcomes.
- Barnett, W. S., Hustedt, J.T., Friedman, A.H., Boyd, J.S., & Ainsworth, P. (2007). *The state of preschool: 2007 state preschool yearbook*. New Brunswick, NJ: Rutgers University, National Institute for Early Education and Research.
- Barnett, W.S., Jung, K., Friedman-Krauss, A., Frede, E.C., Nores, M., Hustedt, J.T., ... & Daniel-Echols, M. (2018). State Prekindergarten Effects on Early Learning at Kindergarten Entry: An Analysis of Eight State Programs. *AERA Open*, 4(2), 2332858418766291. doi: 10.1177/2332858418766291
- Bronfenbrenner, U., & Morris, P.A. (2007). The bioecological model of human development. In R.M. Lerner (Ed.), *Handbook of Child Psychology* (pp. 793-828). New York, NY: Wiley.
- Brown-Jeffy, S., & Cooper, J.E. (2011). Toward a Conceptual Framework of Culturally Relevant Pedagogy: An Overview of the Conceptual and Theoretical Literature. *Teacher Education Quarterly*, 38(1), 65-84.
- Camilli, G., Vargas, S., Ryan, S., & Barnett, W.S. (2010). Meta-analysis of the effect of early education interventions on cognitive and social development. *Teachers College Record*, 112(3), 579-620.
- Cunha, F., & Heckman, J.J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4), 738-782.
- Currie, J., & Thomas, D. (2000). School quality and the longer-term effects of Head Start. *Journal of Human Resources*, 35(4), 755-774.
- DeNavas-Walt, C., Proctor, B.D., & Lee, C.H. (2006). *Current population reports: Income, poverty, and health insurance coverage in the United States: 2005*. Washington, DC: U.S. Census Bureau.
- Dupéré, V., Leventhal, T., Crosnoe, R., & Dion, É. (2010). Understanding the positive role of neighborhood socioeconomic advantage in achievement: the contribution of the home, child care and school environments. *Developmental Psychology*, 46(5), 1227-1244. doi: 10.1037/a0020211
- Early, D.M., Iruka, I.U., Ritchie, S., Barbarin, O.A., Winn, D.-M. C., Crawford, G.M., ... & Pianta, R.C. (2010). How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, 25(2), 177-193. doi: 10.1016/j.ecresq.2009.10.003
- Edelstein, S., Hahn, H., Isaacs, J., Steele, E., & Steuerle, C.E. (2016). *Kids' Share 2016: Federal Expenditures on Children Through 2015 and Future Projections*. Washington, DC: Urban Institute.
- Friedman-Krauss, A.H., Barnett, W.S., Weisenfield, G.G., Kasmin, R., DiCrecchio, N., & Horowitz, M. (2018). *The State of Preschool 2017: State Preschool Yearbook*. New Brunswick, NJ: Rutgers University, National Institute for Early Education and Research.

- Garces, E., Thomas, D., & Currie, J. (2002). Longer-term effects of Head Start. *The American Economic Review*, 92(4), 999-1012.
- García Coll, C.T., Lamberty, G., Jenkins, R., McAdoo, H.P., Crnic, K., Wasik, B.H., & García, H.V. (1996). An integrative model for the study of developmental competencies in minority children. *Child Development*, 67(5), 1891-1914. doi: 10.1111/j.1467-8624.1996.tb01834.x
- Hanson, J.L., Hair, N., Shen, D.G., Shi, F., Gilmore, J.H., Wolfe, B.L., & Pollak, S.D. (2013). Family poverty affects the rate of human infant brain growth. *PLoS ONE*, 8(12), e80954. doi: 10.1371/journal.pone.0080954
- Heckman, J.J., & Karapakula, G. (2019). *The Perry Preschoolers at late midlife: A study in design-specific inference*. Chicago, IL: University of Chicago.
- Hernandez, D.J., Denton, N.A., & Macartney, S.E. (2008). Children in immigrant families: Looking to America's future. *Social Policy Report*, 22(3), 3-22.
- Humphries, M.L., & Iruka, I.U. (2017). Ring the alarm: Moving from educational gaps to educational opportunities for black students. In I.U. Iruka, S.M. Curenton, & T.R. Durden (Eds.), *African American children in early childhood education: Making the case for policy investments in families, schools, and communities* (15-34). Bingley, UK: Emerald Group Publishing Limited.
- Iruka, I.U., LaForett, D.R., & Odom, E.C. (2012). Examining the validity of the family investment and stress models and relationship to children's school readiness across five cultural groups. *Journal of Family Psychology*, 26(3), 359-370. doi: 10.1037/a0028290
- Jeon, L., Buettner, C.K., & Hur, E. (2014). Family and neighborhood disadvantage, home environment, and children's school readiness. *Journal of Family Psychology*, 28(5), 718-727. doi: 10.1037/fam0000022
- Koball, H., & Jiang, Y. (2018). *Basic Facts about Low-Income Children under 9 Years, 2016*. New York, NY: National Center for Children in Poverty.
- Luby, J., Belden, A., Botteron, K., Marrus, N., Harms, N. P., Babb, C., ... & Barch, D. (2013). The effects of poverty on childhood brain development: The mediating effect of caregiving and stressful life events. *JAMA Pediatrics*, 167(12), 1135-1142. doi: 10.1001/jamapediatrics.2013.3139
- Ma, X., Shen, J., Krenn, H.Y., Yuan, J., & Hu, S. (2015). The role of system alignment in care and education of children from birth to grade 3. *Early Child Development and Care*, 185(7), 1067-1087. doi: 10.1080/03004430.2014.978310
- Massey, D.S., Condran, G.A., & Denton, N.A. (1987). The effect of residential segregation on black social and economic well-being. *Social Forces*, 66(1), 29-56. doi: 10.1093/sf/66.1.29
- Matsuda, M., Lawrence, C., Delgado, R., & Crenshaw, K. (1993). *Words that wound: Critical race theory, assaultive speech, and the first amendment*. Boulder, CO: Westview Press.
- McFarland, J., Hussar, B., Wang, X., Zhang, J., Wang, K., Rathbun, A., ... & Mann, F.B. (2018). *The condition of education 2018 (NCES 2018-144)*. Washington, DC: Institute of Education Sciences.
- McKinsey & Company. (2009). *The economic impact of the achievement gap in America's schools*. Washington, DC: Author.
- McLoyd, V.C., & Wilson, L. (1990). Maternal behavior, social support, and economic conditions as predictors of distress in children. In V. McLoyd & C.A. Flanagan (Eds.), *Economic stress: Effects on family life and child development* (50-67). San Francisco, CA: Jossey-Bass.
- Meyers, M.J., & Jordan, L.P. (2006). Choice and accommodation in parental child care decisions. *Community Development: Journal of the Community Development Society*, 37(2), 53-70.
- Mistry, R.S., Benner, A.D., Biesanz, J.C., Clark, S.L., & Howes, C. (2010). Family and social risk, and parental investments during the early childhood years as predictors of low-income children's school readiness outcomes. *Early Childhood Research Quarterly*, 25(4), 432-449. doi: 10.1016/j.ecresq.2010.01.002

### CHAPTER 3 USING A SOCIAL DETERMINANTS OF EARLY LEARNING FRAMEWORK TO ELIMINATE EDUCATIONAL DISPARITIES AND OPPORTUNITY GAPS

- Pungello, E.P., Kainz, K., Burchinal, M., Wasik, B.H., Sparling, J.J., Ramey, C.T., & Campbell, F. A. (2010). Early educational intervention, early cumulative risk, and the early home environment as predictors of young adult outcomes within a high-risk sample. *Child Development, 81*(1), 410-426. doi: 10.1111/j.1467-8624.2009.01403.x
- Raver, C.C., Gershoff, E.T., & Aber, J.L. (2007). Testing equivalence of mediating models of income, parenting, and school readiness for white, black, and Hispanic children in a national sample. *Child Development, 78*(1), 96-115. doi: 10.1111/j.1467-8624.2007.00987.x
- 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. doi: 10.1177/2332858416657343
- Reid, J.L., Kagan, S.L., Hilton, M., & Potter, H. (2015). *A better start: Why classroom diversity matters in early education*. Washington, DC: Poverty & Race Research Action Council and The Century Foundation. Retrieved from <https://files.eric.ed.gov/fulltext/ED571023.pdf>
- Reynolds, A.J., Temple, J.A., Ou, S.-R., Arteaga, I.A., & White, B.A.B. (2011). School-based early childhood education and age-28 well-being: Effects by timing, dosage, and subgroups. *Science, 333*(6040). doi: 10.1126/science.1203618
- Shonkoff, J. P., Garner, A. S., The Committee on Psychosocial Aspects of Child Family Health, C. o. E. C., Adoption, Dependent Care, Section on Developmental Behavioral Pediatrics, Siegel, B. S., Dobbins, M. I., Earls, M. F., . . . Wood, D. L. (2012). The lifelong effects of early childhood adversity and toxic stress. *Pediatrics, 129*(1), e232-e246. doi:10.1542/peds.2011-2663
- Shonkoff, J. P., & Phillips, D. (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press.
- Smith, I. E. (2016). *Minority vs. minoritized: Why the noun just don't cut it*. Retrieved November 16, 2019 from <https://www.theodysseyonline.com/minority-vs-minoritize>.
- Tout, K., Magnuson, K., Lipscomb, S., Karoly, L., Starr, R., Quick, H., ... & Wenner, J. (2017). *Validation of the quality ratings used in quality rating and improvement systems (QRIS): A synthesis of state studies*. OPRE Report #2017-92. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.
- U.S. Department of Health and Human Services. (2010). *Head Start impact study. Final Report*. Washington, DC: Author.
- Urban Institute. (2019). *Segregated from the start: Comparing segregation in early childhood and K-12 education*. Washington, DC: Author.
- Vandell, D.L., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N., & NICHD Early Child Care Research Network. (2010). Do effects of early child care extend to age 15 years? Results from the NICHD Study of Early Child Care and Youth Development. *Child Development, 81*(3), 737-756. doi: 10.1111/j.1467-8624.2010.01431.x
- Whitebook, M., McLean, C., Austin, L.J.E., & Edwards, B. (2018). *Early Childhood Workforce Index*. Berkeley, CA: University of California, Berkeley, Center for the Study of Child Care Employment.
- Whitehurst, G. J. (January 16, 2013). Can we be hard-headed about preschool? A look at Head Start. Brown Center Chalkboard Series. Retrieved from <https://www.brookings.edu/research/can-we-be-hard-headed-about-preschool-a-look-at-head-start/>
- Zellman, G.L., & Perlman, M. (2008). *Child-care quality rating and improvement systems in five pioneer states*. Santa Monica, CA: RAND Corporation.



**SECTION 2**

**WHAT STILL NEEDS  
TO BE UNDERSTOOD?**

## **IN SECTION 2:**

---

### **Chapter 4: Making Prekindergarten Classrooms Better Places for Children’s Development.**

By Dale C. Farran, Ph.D., Vanderbilt University

### **Chapter 5: Improving Quality and Impact Through Workforce Development and Implementation Systems.**

By Robert C. Pianta, Ph.D., University of Virginia and Bridget K. Hamre, Ph.D., Teachstone

### **Chapter 6: Addressing Equity in the ECE Classroom: Equal Access and High Quality for Dual Language Learners.**

By Linda M. Espinosa, Ph.D., University of Missouri-Columbia

### **Chapter 7: Vignette: Building a High-Quality Program—the Boston Public Schools Experience.**

By Jason Sachs, Ed.D., Boston Public Schools

**SECTION 2, CHAPTER 4**

# MAKING PREKINDERGARTEN CLASSROOMS BETTER PLACES FOR CHILDREN'S DEVELOPMENT

---

Dale C. Farran, Ph.D., Vanderbilt University



In this paper, I review four classroom elements that my own work and many other studies have found to be related positively to children's outcomes in prekindergarten classrooms: teachers' listening to children, quality of instruction, emotional climate in the classroom, and level of children's engagement. These aspects of classroom functioning all involve interactions between children and teachers, and they are somewhat independent of both the curriculum and other structural features of the classroom. We need to develop practical observational tools to assess these behaviors so that we can improve the environments in which vulnerable young children learn.

In 2016, according to the National Center for Education Statistics, 66% of 4-year-old U.S. children not in kindergarten were enrolled in pre-primary programs. As in years past, higher-income families were more likely than lower-income families to enroll their children in center-based care. Children from higher-income families often attend privately operated center-based child care programs, while children from lower-income families are likely to be enrolled in publicly funded programs such as Head Start and, more recently, state-funded prekindergarten programs (McFarland et al., 2017).

One consequence of this division is that segregation of experiences by income begins in preschool. Moreover, privately and publicly funded programs have very different expectations and regulations. The fundamental motivation for the two sets of programs differs as well: private child care programs are more concerned with "care" and being of service to parents, while the public programs are more concerned with compensatory education to remediate presumed deficits in children's preparation for school. This desire to offer compensatory education can lead to a greater emphasis on academic preparation and to more prekindergarten programs in public schools. An academic emphasis can have the unfortunate consequence of increased reliance on the sort of didactic instruction that may not lead to long-term child success (Lipsey, Farran, & Durkin, 2018).

## COMPENSATION ORIENTATION

---

Beginning in 1965 with Head Start, a number of public programs for young children before formal school entry focused on compensatory education (Farran, 2007; Scarr & Weinberg, 1986). This trend continued with a 1987 amendment to the Elementary and Secondary Education Act that allowed Title I funds to be used for whole-school program improvement, ushering in the creation of Title I-funded prekindergarten classes in many school districts (Ewen, Mezey, & Matthews, 2005). Although they are administered through different agencies, Head Start and Title I are similar in that neither was intended to provide full-day care; they usually operate on the same schedule as public schools. Although some programs offer before- or after-school care that working families may need, many do not.

Over time, as many states either have begun providing state funds for early intervention prekindergarten programs for children from low-income families or have started to coordinate sources of funding for these programs at the state level, the number of children served has increased. In 2016, most states were funding prekindergarten programs, and a few were offering universal prekindergarten for all 4-year-olds (Barnett et al., 2017). These state-funded programs are primarily intended as compensatory education for children from poor families; all but a few of the states have income requirements for enrollment.

An ethical commitment to using education to remediate or prevent the effects of poverty was put into action in the late 1950s and early 1960s with a number of small experimental programs focused on young children from poor families (Darlington, Royce, Snipper, Murray, & Lazar, 1980). A belief in the efficacy of early education intervention remains a driving force behind the growth in prekindergarten programs (see Parker, Workman, & Atchison, 2016), as more recent data indicate that poverty is still associated with long-term poor school outcomes starting at kindergarten entry (Reardon, 2011).

Since their inception, however, the long-term effectiveness of these small experimental pre-school programs has been debated. Four decades ago, Darlington, Lazar, and others recruited eight of these early experimental programs, including the Perry Preschool Project, and organized a follow-up investigation of their effects (Darlington et al., 1980). The results of their work continue to shape expectations for prekindergarten programs today. They found that the large effects seen on tests given immediately after the programs faded over the next three to four years. However, they found longer-term effects on what they termed "meeting the requirements of school"; that is, students from these programs avoided both special education placement and grade retention at higher rates than did students who had not participated in such programs; the reduction in special education placement was the more robust finding. Expectations of decreased retention and lower use of special education services are featured in such current initiatives as Pay for Success, a program seeking private investment in prekindergarten programs (Isaacs, Massey, & Kreeger, 2016).

Perry Preschool, which began in 1962, is now referred to as a model. The other model frequently cited as evidence for the positive effects of prekindergarten is the Abecedarian program, which began in 1972. The long-term effects from these two programs are the ones most often cited to argue that cost savings will result from extensive investments in preschool (e.g., Heckman, 2006). Both programs served a small number of African American children from low-income families in a single location. Neither has been implemented in any version of a scaled-up statewide program. Each would cost much more per child than any state currently allocates. In today's dollars, Perry would cost \$20,000 per child per year, and Abecedarian would cost between \$16,000 and \$40,000 (Minervino & Pianta, 2014). Moreover, these programs had features that are unlikely to be duplicated. For example, Abecedarian began when children were 6 weeks old, continued until kindergarten, and provided full-day care for 50 weeks of the year; Perry had a 1:7 teacher-child ratio and required that teachers conduct 90-minute weekly visits with families.

These model programs are also hard to replicate because it is not clear which of their components led to the effects. The most robust long-term outcomes for Abecedarian were positive health effects once the children became adults (Conti, Heckman, & Pinto, 2017). This is not surprising, given that two pediatricians and two nurse practitioners were housed in the same building as the preschool, on the same floor as the infant and toddler classrooms. We know less about other components of the treatment offered by the model programs. The HighScope curriculum emerged from the Perry Preschool program but was not solidified until some years after Perry was implemented (Weikart, 2004). Many of the early programs followed a general enrichment philosophy, providing an environment with lots of materials and caring adults. Abecedarian was a pioneer in group care for infants and toddlers, and the staff created a set of activities for teachers to follow with the youngest children (Sparling & Lewis, 1979).

Even when programs are well defined, have a coherent vision, and have more recent evidence of effectiveness, there are problems taking them to scale (Granger, 2011). In the case of statewide prekindergarten programs, for example, states are trying to scale up an idea, not a well-tested practice (Mitchell, 2001). The idea is that an intervention provided to poor children before kindergarten entry will change their developmental trajectories in major, positive ways, both immediately and into adulthood. Less well defined are the exact processes through which that intervention should be carried out.

**Even when programs are well defined, have a coherent vision, and have more recent evidence of effectiveness, there are problems taking them to scale.**

Having the goal of helping children from poor families be successful in school does not really constitute a vision for prekindergarten program practices (Farran, 2017). All states and the District of Columbia have adopted early learning standards for their state-funded prekindergarten programs (DeBruin-Parecki & Slutzky, 2016). These standards are meant to create a bridge between the prekindergarten and the K-12 system, driving and focusing instruction. Learning outcomes can be achieved in a variety of ways, and the standards do not dictate specific practices. States typically set other general requirements for districts that receive state funding to run prekindergarten classrooms. They must meet a certain adult-child ratio, implement a curriculum chosen from an array of possibilities, provide meals for the children and, in some states, provide a certain number of "hours of instruction." These types of requirements are known as structural features; I will review them next along with alternative indicators.

## CLASSROOM QUALITY INDICATORS

### ► Structural characteristics

Programmatic structural characteristics are the easiest to regulate and monitor, and this is where child care quality rating and improvement systems, Head Start programs, and publicly funded state prekindergarten programs overlap the most. Benchmarks specified by the National Institute for Early Education Research (NIEER), which many states use in expanding state funded prekindergarten programs, have historically emphasized these regulatory features. None of these benchmarks—for example, level of teacher education and number of formal degrees—relates to child outcomes either collectively or separately (Early et al., 2007; Mashburn et al., 2008). A recent thorough investigation of credentialing and early childhood education coursework for teachers (Lin and Magnuson, 2018) found negative effects on classroom quality and child outcomes if teachers had only a high school degree and no early childhood education coursework. However, they found no variation in quality linked to the higher end of preparation—that is, having a bachelor's degree and taking many early childhood education courses. Belief in teacher preparation as a key to providing better classrooms with better outcomes persists, however; new Head Start regulations specified in the 2007 reauthorization of the program required that at least 50% of all

Head Start teachers have a bachelor's degree by 2013. Many but not all state prekindergarten programs require a teacher to have a B.A. and to be certified.

**Without measurable alternatives, scaled-up early childhood programs have little guidance for creating quality classrooms, despite calls for the early childhood policy field to focus more on increasing quality.**

What makes these structural characteristics so appealing to law- and policymakers is that they are concrete and measurable: for example, if the rule is a 10:1 child to teacher ratio for 4-year-olds' classrooms, programs can implement that and regulators can check on it. Even though these features are unrelated to children's outcomes, without measurable alternatives, scaled-up early childhood programs have little guidance for creating quality classrooms, despite calls for the early childhood policy field to focus more on increasing quality (Hamre, 2014).

### ► Early childhood curricula

Many quality rating and improvement systems, the NIEER benchmarks, and state-funded prekindergarten programs require a specified curriculum. Many states have lists of curricula that programs can choose; they range greatly both in content and pedagogical strategies (Farran & Lipsey, 2016). When it established its Preschool Curriculum Evaluation Research Consortium in 2001, the Institute of Education Sciences energized the belief that curricula could encompass both the content to be taught and the approaches to learning important for children's growth.

This large experimental research endeavor found few differences in children's outcomes among the curricula assessed or between using a formal curriculum and conducting early childhood classrooms as usual (Preschool Curriculum Evaluation Research Consortium, 2008). The few short-term differences found were positive effects for more targeted curricula—specifically for the outcomes on which they were focused.

Researchers continue to assert the relative advantage of a targeted curriculum over a more global one (the latter often termed “developmental”) (Coley, Votruba-Drzal, Collins & Cook, 2016; Nguyen, 2016). However, even targeted curricular approaches often fail to demonstrate effectiveness. The recent large-scale randomized controlled trial of the Building Blocks preschool mathematics curriculum in New York City found few positive effects compared to control classrooms at the end of the prekindergarten year (Mattera, Jacob, & Morris, 2018). Similarly, another comprehensive review identified few targeted approaches with positive effects that lasted into kindergarten (Chambers Cheung, & Slavin, 2016).

One reason that such curricula may have only short-term effects on the skills they target is that they do not change more fundamental classroom practices. Though teachers may conduct very different activities, as with the Tools of the Mind curriculum, their interactions with their students, the amount of positive feedback they give, and even the amount of time they spend talking and listening to children may be equivalent across different curricula (Nesbitt, Farran, & Fuhs, 2015). Importantly, those interactive elements are the classroom practices linked to children's outcomes in various domains and across curriculum conditions. By itself, no curriculum is likely to effectively or sufficiently drive the kinds of classroom practices that matter most for young children.

### ► Process characteristics

So what should early childhood classrooms, especially scaled-up prekindergarten classrooms, focus on to encourage quality learning? Burchinal reviews global ratings of classroom practices elsewhere in this volume; her research and several other reviews have consistently found little relation between global measures of classroom quality and how children develop over the prekindergarten year. Experimental and descriptive work is currently being done in prekindergarten classrooms to identify more specific behavioral practices as an alternative to such global ratings (e.g., Farran, Meador, Christopher, Nesbitt, & Billbrey, 2017). Many of the practices identified are components of such global instruments as the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008), but this new research disaggregates them from an overall rating of a dimension. Moreover, these approaches are often counts of certain behaviors rather than ratings. A record of the frequency of actual behaviors may offer coaches a clearer way to understand how to help teachers improve their practices.

The work reported by my colleagues and I (Farran, Meador, Christopher, Nesbitt, & Billbrey, 2017) is the result of my four-year partnership among myself, a group of researchers in the Peabody Research Institute at Vanderbilt

University, and the Metro Nashville Public Schools. This work derived from an observation system developed for research purposes in the 1990s (Farran, Silveri, & Culp, 1991). Highly trained and reliable observers remained in classrooms for a full day, taking data throughout the day, several times a year. The system yielded important information about practices that mattered most for young children's growth over the year and even into kindergarten and first grade. The practices determined to be important for children's growth over the preschool year came to be called "the Magic 8" by teachers and coaches in the school system. The appendix contains an example of how one of the practices, reducing transitions, was translated into a tool for coaches to use in our continuing partnership with the district.

Four areas among the eight—teachers' listening to children, quality of instruction, positive climate, and child engagement—have also been investigated and found promising in several other studies.

*Teachers' listening to children matters more than their talking to them.* Language development, and specifically vocabulary, has been one of the hardest areas to improve in early childhood classrooms. In general, however, few links have been found between teacher talk and child outcomes. Our research has shown that the amount of time teachers spent listening to children was actually the stronger predictor of children's growth. While our various studies involving observations of teacher talk show that teachers routinely talk 70% of the time on average, and some talk even more, they spend only about 14% of their time listening to children, on average. Variations in that proportion were important—the more listening teachers did, the more children gained in both academic and social domains. Interestingly, in Dickinson and Porche's (2011) longitudinal study from prekindergarten to fourth grade, it was the ratio of teacher talk to child talk during free play that related to positive outcomes for both kindergarteners and fourth graders. A more even ratio indicated more actual conversations, in which teachers listened to children as well as talked.

In a very complex analysis of the linguistic environment in prekindergarten classrooms (Justice, Jiang, & Strasser, 2018), teachers' linguistic responsiveness—specifically, their facilitation of children's communication—was the only language dimension associated with children's gains in vocabulary. CLASS ratings, also collected, were not related to child outcomes. Justice and colleagues (2018) concluded that rather than trying to improve the global nature of a preschool classroom through such measures as CLASS, "professional development efforts provided to early educators should focus most intensively on helping them to both elevate and execute the precise, proximal behaviors that serve to engage children in productive conversations" (p. 89). They used transcripts of interactions with children to describe many dimensions of teacher language; their analysis indicated that teacher language, including grammatical complexity and linguistic diversity, was not related to children's gains across the year. Only the teachers' verbal interactions with and encouragement of children's language contributions mattered.

One issue with investigating the effects of teacher language may be the emphasis on teacher talk. Most research focuses on analyzing components of teacher language such as the richness and type of language the teacher uses in such activities as book reading. Books and literature constitute one obvious way to introduce varied and more complex vocabulary to children. Thus, many researchers have devoted considerable effort to investigating whether various strategies for book reading might be an effective mechanism for effecting gains in children's language development. One thorough review of book reading's effects concluded that the variation among the studies was too great to yield many recommendations for practice (Wasik, Hindman, & Snell, 2016).

*The teacher's quality of instruction is as important as the student's acquisition of basic skills.*

"Productive conversations," especially teachers' asking questions and listening to children's answers, are components of a more general factor related to the quality of instruction. In a recent book, William Gormley (2017) makes a persuasive argument that encouraging critical thinking through inferential teacher-student interactions may be one of the most important experiences in helping children be successful. He also argues that children from disadvantaged backgrounds are less likely to have these kinds of experiences.

An extensive reanalysis of data from the State-Wide Early Education Programs Study and the National Center for Early Development and Learning Multi-State Study of Prekindergarten found that children from poor families were more likely to experience didactic teaching in prekindergarten classrooms (Valentino, 2017). Didactic teaching is characterized by "known-answer" questions, or "basic concepts" (Farran et al., 2017), such as "What color is this?" and "What letter is this?" Valentino (2017) has suggested that "while there is some evidence that directive instruction could actually improve achievement and narrow achievement gaps in the short term . . . , it is arguable that such an approach is still unfavorable in the long term" (p. 29). Indeed, results from a randomized controlled trial evaluation of the statewide Tennessee Voluntary Prekindergarten program support this hypothesis; despite significantly improved achievement upon entering kindergarten, by the third grade, children who had attended prekindergarten programs, primarily in the public schools, were performing less well than children who had not attended (Lipsey et al., 2018).

Quality of instruction has proven extremely difficult to change; in our four-year study, we were unable to change the level of instruction beyond an average of 1.9 on a 4-point scale. Our observational coding system, Teacher Observation in Preschool (TOP, Bilbrey, Vorhaus & Farran, 2007, revised in 2014) records instances of teacher instruction, defined in early childhood settings as any time teachers are engaged with children around a learning focus. In an early childhood classroom, this could include singing songs and helping with pasting and gluing, as well as reading books and practicing counting, among other activities.

When the teachers' task was coded as "instruction," the instructional level was rated on a scale from 1 to 4. Our definition of instructional quality is derived from research conducted by Tizard and colleagues (1980) and confirmed by classroom observations reported by Durden and Dangel (2008) and Hayes and Matusov (2005). A rating of 1 meant that a teacher was working with materials but not specifically teaching content (e.g., sprinkling glitter); a rating of 2 indicated basic skills instruction (e.g., "What color is the glitter?"); a rating of 3 indicated some inferential instruction, with the teacher asking at least one open-ended question ("This glue is sticky. What else is sticky?"); and a rating of 4 indicated a high degree of inferential instruction, in which the teacher used open-ended questions to sustain focus on a topic that resulted in several conversational turns between teacher and children (a discussion of multiple sticky things). Hayes and Matusov (2005) similarly defined conversational partnerships—our levels 3 and 4—as verbal exchanges of genuine inquiries, where the teacher does not know the answer ahead of time. They found these types of exchanges to be rare in classrooms for young children.

**For more inferential (higher-quality) interactions to take place, teachers have to create interesting learning activities that stimulate children's thinking.**

The rating of 1.9—which we found in all four years of our partnership work with 26 classrooms—is characteristic of instruction at a basic skills level. For more inferential (higher-quality) interactions to take place, teachers have to create interesting learning activities that stimulate children's thinking. They have to interact with children for longer than one

conversational round, and they have to be genuinely interested in the sense that children are making of the world (Durden & Dangel, 2008). These kinds of interactions are difficult if not impossible to carry out during whole group instruction, a common pedagogical practice in these classrooms, and teachers were not observed using center times or small groups as opportunities to initiate higher-level instructional interactions. From their observations in similar classrooms, Darden and Dangel concluded:

*When the kind of activity is (a) guided rather than directed by the teacher, (b) authentic, and (c) exploratory, then the teachers' language changes. In these circumstances, the teacher's language (a) is more open-ended, (b) uses higher cognitive demands, and (c) includes functions such as encouraging thinking, making the nature of the conversation more child-initiated, reciprocal and genuine (p. 261).*

Unfortunately, it would be difficult to help teachers create these kinds of authentic learning opportunities in many of the early childhood classrooms we have observed. Perhaps teachers interpret the increased focus on academic preparedness for kindergarten to mean that they should continually and specifically direct student learning. Engaging children in open-ended inquiry might seem counterproductive to the school readiness goal. In our partnership, we made little progress in this area despite working on it for four years.

*Positive classroom climates promote learning, and the importance of a positive learning environment cannot be overestimated, especially for young, vulnerable children who may be having their first educational experience in a formal setting.* The classroom climate is particularly important for at-risk children, who typically have had a higher than average number of adverse childhood experiences. To promote resiliency in such children, the classroom must promote a sense of belonging, with caring and nurturing adults (Sciaraffa, Zeanah, & Zeanah, 2017). A highly negative classroom can actually function as an additional adverse experience, contributing to rather than buffering the cumulative stress that results in long-term negative health and social outcomes.

Barbara Fredrickson's broaden-and-build theory of positive emotions asserts that a mindset broadened by positive approvals is linked to discovery—"discovery of new knowledge, new alliances, and new skills" (2013, p. 815)—the kinds of discoveries likely to be important for longer-term school success. Harsh, demanding environments can lead to increased immediate learning of concrete skills but not to the fostering of connections among ideas or to the delight in solving problems that are so important for learning in depth. In early childhood classrooms, children are also developing expectations for what being a student means and how learning occurs, and those expectations can color their attitudes toward school for years.

Other studies have also shown that a positive emotional climate is an important contributor to children's growth, especially their social-emotional development. At the end of prekindergarten, children who had been in classrooms with the "warmest" profile were rated the most socially competent (Curby et al., 2009). In a study of 60 prekindergarten classrooms in Tennessee and North Carolina, more teacher approvals, fewer disapprovals, and a more positive teacher emotional tone were collectively related to gains in children's self-regulation (executive function) skills over the prekindergarten year (Fuhs, Farran, & Turner, 2013). In our partnership observations, the same constellation of behaviors was linked to children's gains in academic areas as well. In more positive classrooms, children learned more across the year (Farran et al., 2017).

Recent neurological investigations of brain development in young children from differing socioeconomic backgrounds have found early and alarming differences among children from high-poverty backgrounds in brain regions related to language, memory, executive functioning and socioemotional processing (Ursache & Noble, 2016). These differences were apparent at three years of age. Ursache and Noble (2016) have posited that a causal factor is the stress young children experience in low-income families and neighborhoods. Experiencing frequent disapproval of their behavior in the classroom adds to that stress. In a study of the emotional climate in 139 classrooms serving children from low-income families, recently funded by the Preschool Development Grant Expansion, Durkin and I found high rates of

behavior disapproval—about three times the rate of approval (Farran and Durkin, 2017). Disapproval was especially frequent in classrooms in older public-school buildings without close bathroom or meal facilities. In those types of facilities, the amount of time spent in transitions or down time was greatly increased, and more time in transition was linked to more negative behavioral control (Farran and Durkin, 2017).

The effects of a positive or negative prekindergarten classroom extend into the early grades of school. Two longitudinal studies have demonstrated that the emotional climate of the prekindergarten classroom affects children's social skills into kindergarten and first grade (Broekhuizen et al., 2016; Spivak & Farran, 2016). Reducing behavior disapproval and increasing positive interactions will likely require intense coaching and intervention, as the levels of disapproval are currently quite high in most public prekindergarten classrooms.

*Children's active engagement in learning is key, and engagement should not be confused with compliance.* Children can be quiet and nondisruptive without being engaged. When children are actively involved in learning, they can be noisy (in a productive way). When young children are engaged, they are excited and highly attentive to the learning activity. Engagement is intertwined with all the other components described so far. For example, the level of positive emotional support in a classroom predicted children's level of classroom engagement (Castro, Granlund, & Almqvist, 2017).

Children's active engagement varies across classroom activities. When my colleagues and I observed children in the 26 prekindergarten partnership classrooms (Farran et al., 2017), we found a generally low level of engagement, particularly during whole group instruction. Greater engagement was observed during center-based activities. These findings were echoed in a study of Portuguese pre-schools that also served low-income children; engagement (or involvement) in learning was relatively low for all children (Coelho & Pinto, 2016). Powell and colleagues (2008) carried out extensive research on children's involvement in learning in an "eco-behavioral" investigation. Children were most engaged when teachers were positively affirming and children were with a peer group; they were least engaged during whole group instruction. Vitiello and colleagues (2012) found similar associations between context and child engagement; children were more engaged in situations that gave them some choice over their activities and learning processes.

These findings are important because children in prekindergarten classrooms spend quite a lot of time in whole group instruction and other activities such as transitions where they are under the direction of the teacher. The academic and basic skills orientation of a classroom is linked to greater reliance on whole group instruction and much less to discovery learning. Yet discovery learning is most likely to engage children's attention and keep them focused and involved. Setting up situations where children can be productively engaged in interesting activities requires teachers to act differently as well as to abandon their current understanding of learning.

## CONCLUSION

---

Only recently has public education expanded to offer classrooms for 4-year-olds (McCabe & Sipple, 2011), often housed in public elementary schools. This extension of public education into the prekindergarten years for children from low-income families means that for many children, early childhood settings are now their first introduction to the world of more formal learning and to learning in a group. These early experiences are critical for establishing learning and dispositional patterns that may affect children's interactions with classrooms for many years. The transition to more public school prekindergarten classrooms has happened at the same time that the goal of kindergarten readiness has increasingly come to mean a focus on the mastery of concrete basic skills. Those skills are the very ones most likely to fade quickly in importance in the early grades (Bailey et al., 2016). The kinds of practices outlined in this chapter should be linked both to the mastery of basic skills and to developing more lasting dispositions to learning that will not fade.

In most states, scaled-up publicly funded prekindergarten programs target children from low-income families. Targeting has the unintended consequence of segregating children by income and often by race in their earliest school experience. School districts face a dilemma. They want to place prekindergarten classrooms where the need is—in neighborhoods with a high proportion of poor families and also in underperforming, high-poverty schools—because they believe that better prekindergarten preparation will help children succeed. Such classrooms, housed in buildings not set up for young learners, may then be highly stressful for both teachers and children, leading to more difficult interactions for the children (Gilliam & Reyes, 2018) and unanticipated long-term negative effects on later learning (Lipsey et al., 2018).

Recently, prekindergarten programs have begun moving away from a reliance on regulatory structural features to an emphasis on classroom processes. Yet we lack reliable, easily administered, valid measures for assessing classroom process quality. Many of the quality rating systems that states use, as well as those of current Head Start regulations, include a requirement that classrooms be observed with a rating system like the revised Early Childhood Environment Rating Scale or CLASS. These ratings can be consequential, causing Head Start programs to have to compete again and determining the number of "stars" a private program will receive in a state evaluation. Unfortunately, neither of the most commonly used systems has been shown to predict children's academic or social-emotional growth (Burchinal, 2017).

More recent efforts have focused on specifying the types of classroom interactions that are likely to be most important for children, primarily through behavioral counts instead of ratings. Those efforts have been reported here. They have led to the identification of a number of specific classroom practices that are beneficial for children's learning. The observational system used in the research is not easily exported for use by coaches, principals, or prekindergarten directors. It is complex and requires extensive training. However, the findings can be used to construct more practical and easy-to-use measures. Advances in the digital age should facilitate the collection of critical classroom information. As prekindergarten programs expand, it will become increasingly important to have a system that is practical and can be readily used by coaches, early childhood directors, and principals to assure that children's experiences in these settings are positive and likely to produce long-term benefits.

## References

---

- Bailey, D. H., Nguyen, T., Jenkins, J. M., Domina, T., Clements, D. H., & Sarama, J. S. (2016). Fadeout in an early mathematics intervention: Constraining content or preexisting differences? *Developmental Psychology*, *52*(9), 1457-69. <https://doi.org/10.1037/dev0000188>
- Barnett, S., Friedman-Krauss, A., Gomez, R., Weisenfeld, G., Horowitz, M., Kasmin, R., & Squires, J. (2017). *The state of preschool 2016: State preschool yearbook*. New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.
- Bilbrey, C., Vorhaus, E., & Farran, D. (2007, revised, 2014). Teacher Observation in Preschool [Unpublished scale]. <https://my.vanderbilt.edu/mnpspartnership/research-resources/methods-and-measures>
- Broekhuizen, M., Mokrova, I., Burchinal, M., Garrett-Peters, P., & Family Life Project Key Investigators. (2016). Classroom quality at pre-kindergarten and kindergarten and children's social skills and behavior problems. *Early Childhood Research Quarterly*, *36*(3), 212-222. <https://doi.org/10.1016/j.ecresq.2016.01.005>
- Burchinal, M. (2017, April). Challenges in using widely used observational quality ratings. In M. Burchinal (Chair), *Measuring quality in early childhood education: Issues and promising new instruments* [Symposium]. Biennial Conference of the Society for Research in Child Development, Austin, TX.
- Castro, S., Granlund, M., & Almqvist, L. (2017). The relationship between classroom quality-related variables and engagement levels in Swedish preschool classrooms: A longitudinal study. *European Early Childhood Education Research Journal*, *25*(1), 122-135. <http://dx.doi.org/10.1080/1350293X.2015.1102413>
- Chambers, B., Cheung, A., & Slavin, R. (2016). Literacy and language outcomes of comprehensive and developmental-constructivist approaches to early childhood education: A systematic review. *Educational Research Review*, *18*, 88-111. <https://doi.org/10.1016/j.edurev.2016.03.003>
- Coelho, V., & Pinto, A. I. (2016, June 8-10). Child engagement and interactive processes in inclusive preschool settings [Paper presentation]. International Society on Early Intervention 2016 Conference, Stockholm, Sweden.
- Coley, R., Votruba-Drzal, E., Collins, M., & Cook, K. (2016). Comparing public, private, and informal preschool programs in a national sample of low-income children. *Early Childhood Research Quarterly*, *36*(2), 91-105. <https://doi.org/10.1016/j.ecresq.2015.11.002>
- Conti, G., Heckman, J., & Pinto, R. (2016). The effects of two influential early childhood interventions on health and healthy behavior. *The Economic Journal*, *126*(596), F28-F65. <https://doi.org/10.1111/econj.12420>
- Curby, T., LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., . . . Barbarin, O. (2009). The relations of observed prekindergarten classroom quality profiles to children's achievement and social competence. *Early Education and Development*, *20*(2), 346-372.
- Darlington, R., Royce, J., Snipper, A., Murray, H., & Lazar, I. (1980). Preschool programs and later school competence of children from low-income families. *Science*, *208*(4440), 202-204.
- DeBruin-Parecki, A. & Slutzky, C. (2016). *Exploring prekindergarten age 4 learning standards and their role in early childhood education: Research and policy implications*. Policy Information Report and ETS Research Report Series no. RR-16-14. Princeton, NJ: Educational Testing Service.
- Dickinson, D., & Porche, M. (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. <https://doi.org/10.1111/j.1467-8624.2011.01576.x>
- Durden, T., & Dangel, J. (2008). Teacher-involved conversations with young children during small group activity. *Early Years*, *28*(3), 251-266. <https://doi.org/10.1080/09575140802393793>

## CHAPTER 4 MAKING PREKINDERGARTEN CLASSROOMS BETTER PLACES FOR CHILDREN'S DEVELOPMENT

- Early, D., Maxwell, K., Burchinal, M., Alva, S., Bender, R., Bryant, D., . . . Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78*(2), 558–580. <https://doi.org/10.1111/j.1467-8624.2007.01014.x>
- Ewen, D., Mezey, J., & Matthews, H. (2005). *Missed opportunities? The possibilities and challenges of funding high-quality preschool through Title I of the No Child Left Behind Act*. Washington, DC: Center for Law and Social Policy.
- Farran, D. C. (2007). Is education the way out of poverty? A reflection on the 40th anniversary of Head Start (with commentaries by James King and Bernard L. Charles). *Monographs of the Center for Research on Children's Development and Learning* (University of South Florida), no. 3.
- Farran, D. C. (2017). Characteristics of pre-kindergarten programs that drive positive outcomes. In K. Dodge (Ed.), *The current state of scientific knowledge on pre-kindergarten effects* (pp. 45-50). Durham, NC: Center for Child and Family Policy, Duke University.
- Farran, D. C., & Durkin, K. (2017). Prekindergarten classrooms in the public school buildings: Consequences and concerns [Unpublished manuscript]. Peabody College, Vanderbilt University.
- Farran, D. C., & Lipsey, M. W. (2016). Evidence for the benefits of state prekindergarten programs: Myth & misrepresentation. *Behavioral Science & Policy, 2*(1), 9-18.
- Farran, D. C., Meador, D., Christopher, C., Nesbitt, K., & Billbrey, L. (2017). Data-driven improvement in prekindergarten classrooms: Report from a partnership in an urban district. *Child Development, 88*(5), 1466-1479. <https://doi.org/10.1111/cdev.12906>
- Farran, D. C., Silveri, B., & Culp, A. (1991). Public school preschools and the disadvantaged. In L. Rescorla, M. Hyson & K. Hirsh-Pasek (Eds.), *Early academics: Challenge or pressure?* (pp. 65-73). New Directions for Child Development Source Book. San Francisco: Jossey-Bass.
- Fredrickson, B. (2013). Updated thinking on positivity ratios. *American Psychologist, 68*(9), 814-822. <https://doi.org/10.1037/a0033584>
- Fuhs, M., Farran, D. C., & Turner, K. (2013). Preschool classroom processes as predictors of children's cognitive self-regulation skills development. *School Psychology Quarterly, 28*(4), 347-359. <https://doi.org/10.1037/spq0000031>
- Gormley, W. (2017). *The critical advantage: Developing critical thinking skills in school*. Cambridge, MA: Harvard Education Press.
- Gilliam, W., & Reyes, C. (2018). Teacher decision factors that lead to preschool expulsion: The development and preliminary validation of the Preschool Expulsion Risk Measure. *Infants and Young Children, 31*(2), 93-108. <http://dx.doi.org/10.1097/IYC.0000000000000113>
- Granger, R. C. (2011). The big why: A learning agenda for the scale-up movement. *Pathways* (Winter), 28–32.
- Hamre, B. (2014). Teachers' daily interactions with children: An essential ingredient in effective early childhood programs. *Child Development Perspectives, 8*(4), 223-230. <https://doi.org/10.1111/cdep.12090>
- Hayes, R., & Matusov, E. 2005. Designing for dialogue in place of teacher talk and student silence. *Culture and Psychology, 11*(3), 339-357.
- Heckman, J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*(5782), 1900-1902.
- Isaacs, J., Massey, M., & Kreeger, A. (2016). *Executive summary: Pay for Success early childhood education toolkit*. Washington, DC: Urban Institute. <https://pfs.urban.org/library/ecetoolkit/content/executive-summary-pay-success-early-childhood-education-toolkit>
- Justice, L., Jiang, H., & Strasser, K. (2018). Linguistic environment of preschool classrooms: What dimensions support children's language growth? *Early Childhood Research Quarterly, 42*(1), 79-92. <http://dx.doi.org/10.1016/j.ecresq.2017.09.003>

- Lin, Y.-C., & Magnuson, K. (2018). Classroom quality and children's academic skills in child care centers: Understanding the role of teacher qualifications. *Early Childhood Research Quarterly*, 42(1), 215-227. <http://dx.doi.org/10.1016/j.ecresq.2017.10.003>
- Lipsey, M., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee prekindergarten program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly*, 45(4), 155-176. <https://doi.org/10.1016/j.ecresq.2018.03.005>
- Mashburn, A., Pianta, R., Hamre, B., Downer, J., Barbarin, O., Bryant, D., . . . Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79(3), 732-749.
- Mattera, S., Jacob, R., & Morris, P. (2018). Strengthening children's math skills with enhanced instruction. The impacts of Making Prekindergarten Count and High 5's on kindergarten outcomes. New York: MDRC. [https://www.mdrc.org/sites/default/files/MPC-High\\_5s\\_Impact\\_FR\\_0.pdf](https://www.mdrc.org/sites/default/files/MPC-High_5s_Impact_FR_0.pdf).
- McCabe, L., & Sipple, J. (2011). Colliding worlds: Practical and political tensions of prekindergarten implementation in public schools. *Educational Policy*, 25(1), e1-e26. <https://doi.org/10.1177/0895904810387415>
- McFarland, J., Hussar, B., de Brey, C., Snyder, T., Wang, X., Wilkinson-Flicker, S., . . . Hinz, S. (2017). *The Condition of Education 2017* (NCES 2017-144). U.S. Department of Education. Washington, DC: National Center for Education Statistics. <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2017144>
- Minervino, J., & Pianta, R. (2014, September). Early learning: The new fact base and cost sustainability. In J. Minervino (Ed.), *Lessons from research and the classroom* (pp. 8-20). Seattle, WA: Bill and Melinda Gates Foundation.
- Mitchell, A. (2001, July). Prekindergarten programs in the States: Trends and issues. (ED454948). ERIC. <https://eric.ed.gov/?id=ED454948>
- Nesbitt, K., Farran, D. C., & Fuhs, M. (2015). Executive function skills and academic achievement gains in prekindergarten: Contributions of learning-related behaviors. *Developmental Psychology*, 51(7), 865-878. <http://dx.doi.org/10.1037/dev0000021>
- Nguyen, T. (2016, September). Impacts of targeted and global preschool curricula on children's school readiness: A meta-analytic review (Working paper). University of California, Irvine.
- Parker, E., Workman, E., & Atchison, B. (2016, January). *50 state review: States prekindergarten funding for 2015-16 fiscal year*. Denver, CO: Education Commission of the States.
- Pianta, R.C., La Paro, K., & Hamre, B. (2008). *Classroom Assessment Scoring System (CLASS) manual, prekindergarten*. Baltimore, MD: Paul Brookes.
- Powell, D., Burchinal, M., File, N., & Kontos, S. (2008). An eco-behavioral analysis of children's engagement in urban public school preschool classrooms. *Early Childhood Research Quarterly*, 23(1), 108-123. <https://doi.org/10.1016/j.ecresq.2007.04.001>
- Preschool Curriculum Evaluation Research Consortium (2008). *Effects of preschool curriculum programs on school readiness*. NCER 2008-2009. Washington, DC: U.S. Government Printing Office.
- Reardon, S. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. Duncan & R. Murnane, (Eds), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 91-116). New York: Russell Sage Foundation.
- Scarr, S., & Weinberg, R. (1986). The early childhood enterprise: Care and education of the young. *American Psychologist*, 41(10), 1140-1146.
- Sciaraffa, M., Zeanah, P., & Zeanah, C. (2018). Understanding and promoting resilience in the context of adverse childhood experiences. *Early Childhood Education Journal*, 46(3), 343-353. <https://doi.org/10.1007/s10643-017-0869-3>
- Sparling, J., & Lewis, I. (1979). *Learning games for the first three years: A guide to parent-child play*. New York: Walker.

## CHAPTER 4 MAKING PREKINDERGARTEN CLASSROOMS BETTER PLACES FOR CHILDREN'S DEVELOPMENT

- Spivak, A., & Farran, D. C. (2016). Predicting first graders' social competence from their pre-school classroom interpersonal context. *Early Education and Development, 27*(6), 735-750. <https://doi.org/10.1080/10409289.2016.1138825>
- Tizard, B., Hughes, M., Pinkerton, G., & Carmichael, H. (1982). Adults' cognitive demands at home and at nursery school. *Journal of Child Psychology and Psychiatry, 23*(2), 105-116. <https://doi.org/10.1111/j.1469-7610.1982.tb00056.x>
- Ursache, A., & Noble, K. (2016). Neurocognitive development in socioeconomic context: Multiple mechanisms and implications for measuring socioeconomic status. *Psychophysiology, 53*(1), 71-82. <https://doi.org/10.1111/psyp.12547>
- Valentino, R. (2017). Will public prekindergarten really close achievement gaps? Gaps in prekindergarten quality between students and across states. *American Educational Research Journal, 55*(1), 79-116. <https://doi.org/10.3102/0002831217732000>
- Vitiello, V., Booren, L., Downer, J., & Williford, A. (2012). Variation in children's classroom engagement throughout a day in preschool: Relations to classroom and child factors. *Early Childhood Research Quarterly, 27*(2), 210-220. <https://doi.org/10.1016/j.ecresq.2011.08.005>
- Wasik, B., Hindman, A., & Snell, E. (2016). Book reading and vocabulary development: A systematic review. *Early Childhood Research Quarterly, 37*(4), 39-57. <https://dx.doi.org/10.1016/j.ecresq.2016.04.003>
- Weikart, D. (2004). *How High/Scope grew: A memoir*. Ypsilanti, MI: High/Scope Press.

Appendix

**Reducing time spent in transition: What is a transition?**

A "transition" is a prolonged period in which most of the class is not involved in a learning activity.

### Common Types of Transitions

Breaks when **one activity has ended but another has not yet begun.**

Interruptions of activities that result from teachers **gathering materials or correcting behavior.**

Times that children can't begin an activity because they are **awaiting instructions or materials.**

Times that children are **moving to a new location** (i.e., going outside, lining up for restroom breaks)

**Think of the time spent in a classroom as a pie chart in which every moment is accounted for. If a large "slice" of the day is spent transitioning, less time is available for other learning activities.**

**Reducing time spent in transition leads to:**

1. Fewer instances of problem behavior.
2. Higher levels of involvement in learning.
3. **More time available for instruction.**

**Data collected in MNPS Early Learning Center classrooms showed a strong relationship between time spent in instructional activities and children's achievement gains.**

Certain parts of the day are beyond the teacher's control (e.g., how far the class needs to travel to the playground or the cafeteria). Intentional planning of transitions allows you to create routines that accommodate the classroom schedule and student needs.



**Some transition time during the day is both normal and necessary—the goals for reducing transitions should be to:**

1. Decrease the overall “wait time” between activities whenever possible.
2. Incorporate engaging instructional content when a transition is unavoidable.

### Practical Strategies for Teaching Transition Routines

- Take time at the beginning of the school year to establish expectations for moving from one activity to another.
- Revisit these procedures periodically.
- Model appropriate cleanup behaviors.
- Act out a scenario in which you are cleaning up your area while thinking aloud and allowing children to help you problem solve. Try getting parts of the routine wrong on purpose—children will LOVE to “correct” you!

### Troubleshooting Transitions

If you notice things are still not going smoothly, it may be a good idea to play detective! Sit back and watch as your students transition from one activity to the next. What do you notice?

#### Ask yourself:

- Do I spend a lot of time addressing behavior during transitions?
  - Do I unnecessarily spend time redirecting harmless or minor behaviors?
- Do children who finish transitioning first seem bored while they wait for their peers?
- Do we need to reset or review?
- Are there particular transitions that are stressful for me or for my students?
  - Before the transition → Self-care (take a deep breath)
  - During the transition → Try the Practical Strategies to help minimize time spent in transition
  - After the transition → Make mental notes about what worked or didn't work



**SECTION 2, CHAPTER 5**

**IMPROVING QUALITY AND IMPACT  
THROUGH WORKFORCE DEVELOPMENT  
AND IMPLEMENTATION SYSTEMS**

---

Robert C. Pianta, Ph.D., University of Virginia

Bridget K. Hamre, Ph.D., Teachstone



## INTRODUCTION

---

It is evident that the benefits of publicly funded early education and care programs, while significant, are not sufficiently large to close the notable gaps in children’s learning and development at the start of school. Most efforts to increase program benefits have focused on workforce development, whether through traditional teacher preparation in higher education or through professional development (PD) for practicing teachers. Overall, workforce development has demonstrated mixed benefits, on average, for teachers or children (Fukkink & Lont, 2007; Snyder, Hemmeter, Meeker, Kinder, Pasia, & McCloughlin, 2012). But studies also report examples of proven-effective, workforce-focused PD interventions with significant positive impacts on teacher and student performance (e.g., Neuman & Cunningham, 2009; Landry, Swank, & Anthony, 2011; Pianta, Hamre, Downer, et al., 2017). This chapter addresses the gap between proven-effective PD and efforts to deliver PD that has a widespread impact on the workforce and children. Overall, PD is hampered by, among other things, varying standards across the states; less-than-effective coaches; and gaps between how implementation science says PD should work and how it is put to work in practice. When PD is intentional and integrated, it is more likely to be effective and can provide a better unified, quality experience for children across varied settings and teachers. Areas for improvement include ensuring that PD has a clear focus and targets specific outcomes; supporting the PD workforce; providing course-based PD; and using certified PD providers.

## A FRAGMENTED SYSTEM, VARIED WORKFORCE, INEFFECTIVE APPROACHES

---

Early education and care encompasses many programs under a variety of names and auspices for children who have not yet entered kindergarten. They include state-funded pre-K, community preschools, Head Start, and family- and community-based child care. Many children are enrolled in more than one such program at any given time, and most are exposed to multiple forms of programming at different ages. The result is great variation and fragmentation for children, families, programs, and the workforce, which is reflected not only in children’s exposure to multiple programs, but also in the needs of a workforce whose educational qualifications range from high school equivalents to advanced degrees. Providers often see their programs as existing in silos at the same time that their different approaches and resources constitute a whole experience for children, potentially hampering effective child development. We suspect that effective PD, implemented well across the early education system, could create more continuity and value for children, educators, and families. It may be that a more consistent, systemic focus on a few organizing principles that make for effective teaching and PD—child-centric, teacher-child interactions, intentionality, personalization, teacher-parent interaction—could make the education experience more effective across all the settings a young child may traverse.

**Many children are enrolled in more than one such program at any given time, and most are exposed to multiple forms of programming at different ages.**

As just one example of fragmentation, consider that children can expect a stunning level of variation from year to year and setting to setting in even the most basic qualifications of the early education and care workforce. Although 95% of kindergarten teachers have a bachelor's degree, preschool teachers vary widely in their level of training. On average, they receive less training and education than do their elementary school counterparts (Early et al., 2007; Ryan & Whitebook, 2012). Even among teachers in state-funded pre-K programs, minimum requirements can range from a Child Development Associate (CDA) certificate to an associate degree to a bachelor's degree (Barnett et al., 2016). Furthermore, some states require that the two- or four-year degree be in early childhood education (ECE) or child development, whereas others do not specify a field of study. Thus, even in state-funded pre-K programs and kindergarten, which are fairly well regulated, the preparation and qualifications deemed necessary for the workforce vary substantially. Head Start has national standards for program structure, operation, and teacher credentials but does not require all teachers to have college degrees. In 2007, Head Start increased its educational standards for teachers and educational coordinators, such that a minimum of 50% of lead teachers would have at least a bachelor's degree by 2013, a goal that was attained at considerable expense in time, effort, and funds.

For children enrolled in less-regulated family- or center-based child care, exposure to credentialed or degreed staff is even lower (National Registry Alliance, 2013; Ryan & Whitebook, 2012). The National Association for Regulatory Administration's 2008 child care licensing study (NCCITA & NARA, 2010) was one of the most comprehensive examinations of the child care workforce. Data from 49 states and the District of Columbia showed that in the vast majority of states (42), child care-center directors are required to have only some occupational–vocational training, some higher education credit hours in ECE, or a CDA credential. Only one state required that directors hold a bachelor's degree. Similarly, for individuals considered as teachers in licensed child care centers, 40 states required some combination of a high-school degree and experience. Only 10 states required a vocational program, certificate, or CDA, and 13 had no requisite educational qualification for child care teachers—a pattern of low-level qualifications and compensation that remains the case today (Whitebook, Phillips, & Howes, 2014).

Clearly, states (and the field in general) have not settled on a set of minimum qualifications for adults serving as teachers of young children, whether they work in private child care, Head Start, or public pre-K. To the extent that these settings are expected to contribute to children's learning and development, then characterizing these adults as teachers and explicitly outlining qualifications and competencies aligned to that role would be a first step. Moreover, there is little agreement on the performance standards that should be applied to this role or on how to measure those standards, and the preparation and PD experiences that should align with such performance standards are woefully out of synchrony.

Unsurprisingly, given the uncertainty regarding basic qualifications, the variation in the nature and quality of training, and the low compensation for the early education and care workforce (Whitebook et al., 2014)—which

discourages higher education—it’s difficult to provide effective training and PD. Given the increased costs associated with additional training and degrees, it becomes even more important to justify the costs by showing that those experiences impact students’ learning and achievement. We know too little about the knowledge and competencies that representative members of the workforce display and how such knowledge and competencies map to the needs and outcomes of the children they serve, or the focus and impact of curricula or PD programs. And we have good evidence that the early education and care workforce experiences high levels of stress and workplace demands that undermine the quality of the care it provides (Whitaker, Dearth-Wesley, & Gooze, 2014). Moreover, PD and workforce training in early education and care is not often tailored to the individual professional’s needs, or to curricula or programs being implemented; instead, it is fairly generic, loosely coupled to practice, and marginally effective. Overall, the early education and care workforce operates on razor-thin margins of support, whether it be in the form of compensation, regulation, or PD.

## GAPS IN KNOWLEDGE, TOOLS, AND IMPLEMENTATION

The disconnect between the needs of the early childhood workforce and scaled implementation of effective PD is a tremendous impediment to improving young children’s learning. This is true even when a number of early childhood workforce PD models in controlled evaluations have demonstrated benefits for teachers and for children (e.g.,

**The disconnect between the needs of the early childhood workforce and scaled implementation of effective PD is a tremendous impediment to improving young children’s learning.**

Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008; Hamre, Pianta, Burchinal, & Downer, 2012; Raver et al., 2008). One would think that these models, once made available or disseminated, would be adopted and would yield expected benefits, at least according to the logic of research, development, and dissemination that underlies most education science (Pianta & Hofkens, 2018). Yet, even though federal and state funding has poured into initiatives emphasizing support for the teaching workforce—including for Head Start quality improvement and teacher education, for the Race to the Top-Early

Learning Challenge, and for Quality Rating and Improvement Systems—few benefits have been detected for children or for teachers’ skills (Pianta & Hofkens, 2018).

Recognizing the need and value for PD, policymakers have made significant investments in the workforce, which is a first step. But that investment does not focus enough on proven-effective PD models. Unfortunately, teachers rarely experience PD that reflects features of specificity and alignment to practice. In fact, a recent survey that was representative of the 1 million teachers in center-based programs for children aged 0 to 5 years indicates that the predominant form of PD is a one-hour workshop only tangentially connected to teachers’ everyday practice and known to be ineffective (McCormick Center for Early Childhood Leadership, 2016; Zaslow, Tout, Halle, Whittaker, & Lavelle, 2010).

We reflect on this conundrum from the perspective of having worked for more than a decade to develop, evaluate, implement, and scale tools for assessing and improving quality in early education and care through workforce development. Collectively, our activities have included: training observers to acceptable standards of agreement; supporting large-scale implementation of classroom observations; developing and evaluating coaching, coursework, and workshop-style PD; and working with states and local systems' own initiatives that draw from an assortment of tools. This work has ranged from the early stages of research and development to implementation at scale.

In these efforts, we have witnessed stunning variation in state and local needs, workforce strengths, goals for program improvement or child learning, and the skill and knowledge profiles of educators (Barnett et al., 2015). We have also noted the manner in which this variation—at all levels and in all forms—intersects with the goals of standardization, consistency, and fidelity that are paramount in developing, evaluating, and using educational programs and tools to produce the effects for which they are designed and intended. Most of the time, the conditions that render a PD model or tool “proven effective” are misaligned with the realities of local programs and staffs, which constrains the extent to which even the best-developed and easiest-to-use tool fits local needs or goals and can be implemented locally with consistency and potency.

Implementation science can offer a framework for knitting together the potential of proven-effective training and PD with the everyday realities of classroom practice, program capacity, and surrounding systems. This is because implementation science, with its focus on identifying and engineering the conditions that influence and explain strong and weak implementation, can create the kind of systemic and aligned programs of professional training and development that foster improvements in classrooms and impacts on children.

We see tremendous potential for progress. At no other time has the field been as poised to enable sustained, positive change. Multiple stakeholders now recognize that high-impact implementation through redesigned workforce development is the key to making good on investments in access made over many decades. We understand that classroom processes are the key mechanisms through which workforce development transmits benefits to children. Effective tools (curricula, assessments, coaching models) are available. We know more now about workforce needs than we did 10 years ago. And research on the elements of effective PD provides a steady stream of largely consistent findings. Yet, although research has generated considerable new knowledge and a wide range of tools for classroom use, successful translation and use of that knowledge is spotty and weak. The essential gaps, regardless of whether we have evidence-supported tools and curricula, reside in systems for using, applying, and implementing knowledge.

## A STARTING POINT: FEATURES OF PD THAT IMPROVE TEACHING AND LEARNING

Reports have clearly described the features of PD that relate to improved practice and student learning (Zaslow et al., 2010). When targeted, practice-aligned PD supports are available to teachers, student skill gains can be considerable—at times on the order of half a standard deviation and higher in some subgroups. Recent meta-analyses of PD for early childhood educators have shown positive effects at the classroom, educator, and child levels (Markussen-Brown, Juhl, Piasta, Bleses, Hojen, & Justice, 2017). For example, in the social-emotional domain, PD that focuses directly on child care providers' interactions with children leads to higher-quality classroom environments, adult-child interactions, and child behaviors (Werner, Linting, Vermeer, & Van IJzendoorn, 2016). In language and literacy, PD improved teaching and children's phonological awareness and alphabet knowledge (Markussen-Brown et al., 2017). Larger effects are typically reported more for proximal outcomes (e.g., classroom- or teacher-level) than for distal outcomes (e.g., children's learning), a finding that is common in the PD literature more broadly.

### ► Focus on teacher skills and relevant knowledge

A starting point for identifying, implementing, and eventually scaling effective PD is to consider the PD target and the system in which it will be implemented. As Burchinal (this volume) suggests, classroom observation of teacher practice is often viewed as a source of information on the focus or target of PD, as is teachers' knowledge of children's development or of a curriculum. To the extent that such practices or knowledge reflect features of quality that are linked to children's learning, there is a stronger basis for selection as a focus for PD.

Several examples demonstrate the systematic use of validated tools to observe teachers' practice as a focus for PD. For example, Hemmeter, Fox, & Snyder (2013) have used the Teaching Pyramid Observation Tool (TPOT) (Fox, Hemmeter, & Snyder, 2014) to guide their coaching work, which focuses on teachers' support for children's social and emotional skills. The TPOT measures a set of practices, identified in research on classrooms, which are known to promote positive behavior among young children. From the standpoint of linked PD, coaches who use the Practice-Based Coaching approach to intervention conduct TPOT observations to define targets for their work with teachers. Several studies have shown that linking TPOT observations to coaching on specific TPOT-identified and described behaviors leads to changes in teachers' practice (Hemmeter, Fox, & Snyder, 2013; Hemmeter, Hardy, Schnitz, Adams, & Kinder, 2015). Moreover, this approach has been shown to improve children's teacher-reported and -observed social skills, which is the model's desired outcome.

In another example of scaled-up PD linked to targeted observations, Landry, Anthony, Swank, and Monseque-Bailey (2009) built many of their effective coursework and coaching approaches explicitly from the CIRCLE TBRS (Landry, Crawford, Gunnewig, & Swank, 2002), an observational measure articulating 50 specific teaching behaviors that have been linked to children's development and learning in the social-emotional and literacy domains.

PD models designed around the Classroom Assessment Scoring System (CLASS) (Pianta, La Paro, & Hamre, 2008) include a college course and a video-based coaching model that have demonstrated positive impacts on teaching practice and, in some studies, on student outcomes (Downer et al., 2011; Pianta et al., 2017; Pianta, Mashburn, Downer, Hamre, & Justice, 2008). Hamre et al. (2012) demonstrated that the course improved the quality of teachers' interactions with children and their observation skills, an effect that remained detectable a year later (Downer et al., 2011). Experimental evaluations of MyTeachingPartner (MTP) coaching showed improvements in pre-K teachers' interactions with students, effects that doubled in high-poverty classrooms. When teachers received MTP coaching, children made greater gains in receptive vocabulary, task orientation, and prosocial assertiveness. A second evaluation of MTP, using local coaches with 450 pre-K teachers at 15 sites, showed that coaching improved nearly every CLASS dimension (and particularly instructional support), with effect sizes averaging .5 to .75 standard deviations (Downer et al., 2011), and produced overall gains in children's self-regulation skills and classroom-level language behavior (Pianta et al., 2017). In classrooms where children differed little in age, benefits were detected for children's literacy and language development skills as well (Ansari & Pianta, 2018). Notably, there was some evidence of a dose-response relation between the amount and target of MTP coaching and the level and dimension of gain in teachers' quality of interaction (Pianta et al., 2014).

In the area of teaching practices that support children's development in mathematics, Clements and colleagues (Baroody, Clements, & Sarama, 2019) have repeatedly demonstrated an impact on teachers and children from observing teachers' practice, both generally and while implementing a curriculum, and the potency of providing them with feedback, modeling, and coaching support within an integrated curricular and PD package (Clements et al., 2018). And in science education, Piasta, Logan, Pelatti, Capps, and Petrill (2015) report a similar pattern of findings linking observation with PD to drive improvements in practice.

Many PD programs with demonstrated impacts have used other methods to identify teaching practices to focus on (e.g., Piasta et al., 2012; Williford et al., 2017). As just one example, Barton, Fuller, & Schnitz (2016) developed a performance feedback model for pre-service teachers that targeted seven teacher practices for supporting children in inclusive settings. Those practices were derived from careful analysis of the empirical literature and became a focal point for feedback on candidates' emerging competencies.

It may seem obvious that PD should focus on evidence-based teaching practices, but experience and the limited available data suggest that much PD for teachers does not do so. In one review of 256 published studies of ECE PD, only 25% had explicitly focused on teaching practices (Snyder et al., 2012). And the vast majority of practice-focused PD targets more generalized teaching practices, early literacy, and/or social-emotional teaching (Schachter, 2015).

A meta-analysis of language and literacy PD packages found that including any coaching component resulted in significantly better teacher practice ( $d = .68$  with coaching,  $d = .22$  without coaching; Markussen-Brown et al., 2017). In another meta-analysis, Werner et al. (2016) found that programs including individualized follow-up for teachers had significantly larger effect sizes on teacher outcomes than did programs without that type of follow-up. But most early childhood teachers lack access to coaches or follow-up. Based on data from the National Survey of Early Care and Education (Tout, Halle, Datta, & Snow, 2015), only 36% of preschool teachers reported that they had received any coaching, mentoring, or consultation in the past year.

PD research has also examined teachers' knowledge of practice-relevant information. A few studies have systematically tested the effects of a specific course that aims to enhance knowledge of children's skill development, or of curriculum and practice relevant to implementation, with some promising results (Dickinson & Caswell, 2007; Howes, Galinsky, & Kontos, 1998; Kontos et al., 1996; Neuman & Cunningham, 2009). Neuman & Cunningham (2009) demonstrated that a course focused on knowledge and practices related to fostering young children's language and literacy development impacted the observed practices of child-care providers. Examining a course focused on teachers' knowledge of the dimensions of teacher-student interaction and their skills in identifying different features of interaction, Hamre et al. (2012) found positive impacts on teachers' classroom interactions that approached the effects of coaching. And Clements and colleagues (Clements et al., 2018) recently reported that exposing teachers to information on children's learning trajectories can improve practices in mathematics instruction. In sum, the evidence clearly shows that when PD provides selective and practice-relevant information, teacher and child outcomes can improve.

### ► **Ensure sufficient intensity and duration**

Intensity and a greater duration of PD consistently leads to improvements in teachers' practice (Garet, Porter, Desimone, Birman, & Suk Yoon, 2001; Markussen-Brown et al., 2017). Markussen-Brown and colleagues (2017) reported a wide range of intensity among the studies they included in their meta-analysis of PD, from six to 450 total hours; they found greater changes in teaching practice among PD programs with greater intensity. Unfortunately, we do not know exactly how much PD is enough, though it is likely that the answer depends greatly on the desired outcome. Smaller elements of practice can change as a result of relatively moderate-intensity PD. For example, Promoting Early Literacy in Licensed Care (PELLC) was designed to be a modest effort in terms of dosage and cost (Gerde, Duke, Moses, Spybrook, & Shedd, 2014), with a course consisting of five sessions, each lasting two hours, for a total of 10 hours of PD. Evaluation of the PELLC course found significant effects on providers' literacy knowledge and practices, but no evidence of impacts on children's literacy outcomes.

Some compelling studies have systematically varied intensity and duration in ways that provide causal evidence. Landry, Swank, Anthony, and Assel (2011) had teachers participate in nine online workshops and receive in-person mentoring twice a month across the year. Some teachers received the intervention for one year, and others for two years. The researchers found that one year of the intervention had significant effects on teachers' language and literacy instructional practices. A second year of coaching produced no additional impact on teaching practice but had larger impacts on children's learning. It takes some time for teachers to change their practice (Pianta et al., 2014), and it may be that children in teachers' classrooms during the first year of PD would not have enough exposure to the improvements in practice to show demonstrable impact. Systematically varying dosage in research studies could help refine our understanding of how much PD is needed to support specific types of practice changes, and this could be a focus for implementation research.

In sum, ample evidence from rigorous experimental studies shows that PD focused on teacher practices or relevant knowledge can improve the quality of teachers' skill and, to a lesser extent, children's learning. We have curricula, methods of practice, and tools that can predictably improve teachers' knowledge and skill, and a number of them also show evidence of further benefits for children's learning. At the same time, there is fairly broad agreement that PD for ECE teachers as typically implemented by states and school systems throughout the country is not all that effective. The opportunity to deploy PD investments for greater impact holds tremendous promise for improving the benefits of programs for children.

## **SYSTEMS SUPPORTING HIGH-FIDELITY IMPLEMENTATION AND SCALE-UP OF EFFECTIVE PD**

---

To improve the quality and impact of programs at scale through workforce development, we must explicitly specify the enabling architecture—the incentives, standards, training and implementation protocols, quality control procedures, and certifications that shape the actions of various people in the system (teachers, purveyors, programs) to produce high effort and focused participation. All too often, these components of a workforce development system are misaligned with one another, with the needs of the workforce, and with the support structures needed to deliver the types of proven-effective PD described here.

Most of the time, PD requirements are established by state licensing regulations that structure educators' career development (Whitebook, Bellin, Lee, & Sakai, 2005). These regulations are typically generic—for example, the number of PD hours teachers need to complete for licensure renewal. Rarely do regulations specify the target, content, quality, or impact of PD. Most administrators lament relying on "hours accumulated" as the metric for linking PD to an incentive structure because it almost guarantees a lack of focus or alignment to teachers' skill needs or specific areas for curricular or classroom improvements. In this sense, PD is untethered from individual needs for training or local program plans. Even teachers themselves report significant failure in the PD system. When the

McCormick Center for Early Childhood Leadership (2016) surveyed over 500 teachers working across program types (75% with a BA or higher), fewer than half of respondents (43%) believed that their PD opportunities “were very helpful in strengthening their level of professional competence” (pp. 1-2). If many millions of dollars are spent on PD each year (to say nothing of the costs related to paying teachers for hours spent in PD, or the opportunity costs of attending PD that has no impact), and if PD presumably plays a critically important role in advancing the benefits of early care and education, then why are things so broken?

The primary gaps in workforce development involve mechanisms to explicitly integrate knowledge, tools, workforce needs, and incentive structures in a program improvement and workforce development system that enables rigorous and potent implementation of proven-effective approaches and systematic use of data for improvement. Without steady and close integration of two activities—mapping proven-effective PD models into a system for scaling with fidelity—most teachers will attend serial one-time workshops at considerable personal and public cost. These activities and the time teachers spend will have little to no chance of benefitting them or their students.

Let’s look at one example of this interface between a PD model and a scaling system. In a recent implementation of a new QRIS, the state of Louisiana chose to use CLASS as the metric for quality, and hence the sole target for improvement through PD (enabled by incentives). Louisiana then identified a small set of PD models that, in controlled evaluations, had been shown to improve CLASS scores. The state then created systems of incentives aligned to increase teachers’ and programs’ selection of those models—for example, legislation linked tax credits for providers to their engagement with these effective PD models. In addition, higher education programs that prepare teachers with bachelor’s degrees for the state pre-K program would soon need to align their content and assessments to the QRIS targets. Moreover, this move to scale also included procedures for ensuring reliable collection of CLASS scores, training for PD providers, and other enabling features, such as evaluation and quality control analyses. Thus, the approach was both systemic and systematic.

In this illustration, models of PD that had been proven effective in rigorous studies were integrated in a scaling system that drew on the QRIS and tax-credit system as a way to encourage and enable use at a wider scale. The Louisiana example is perhaps a template for scaling up that integrates and aligns systems of large-scale implementation with PD models that have proven potential for impact. Most notably, the Louisiana model reflects an overall strategy and explicit design for a system of inputs to teachers and the enabling infrastructure.

## CONDITIONS FOR IMPLEMENTATION WITH IMPACT

---

We have described promising findings that suggest PD can reliably and confidently produce benefits for teachers and children, as well as the parallel challenges of promoting such proven-effective PD at scale. Next, we identify several conditions that are key to closing the gaps between PD that has been proven effective under local or controlled conditions to implementation with benefits at scale.

### ► Use a clear and focused PD program or model

Zaslow and colleagues (2010) have described the features of effective PD programs, which include a focus on: a) students' skill targets and developmental progressions (e.g., developmental progressions in decoding skills); b) improving teachers' skillful use of instructional and social interactions to promote student engagement and learning (e.g., feedback or conversation); and c) fostering teachers' skills and knowledge to effectively implement curricula and appropriately engage children with content (e.g., delivering an effective and engaging activity on teaching cardinality). These features all emphasize a defined and relevant set of knowledge and practices as enacted by teachers. In recent meta-analyses of PD in ECE (e.g., Markussen-Brown et al., 2017), most of the effective PD models were based on evidence linking focal practices to specific child outcomes. Some effective PD models also focus on teacher knowledge, which, if tightly linked to practice, can make positive changes to teachers' daily work in the classroom. As we note, a number of bundled curricula and PD supports have shown a proven impact on student learning; similarly, we have some examples of PD focused on general teacher practices with known relations to student outcomes. These are the starting places for decisions and investments aimed at scale.

The alignment of PD, curricula, assessment, and other enabling supports creates a sort of operating system for a program, an important factor in success. Most recently, Connors, Pacchiano, Manos, & Horsley (2018) described how the Ounce of Prevention Fund fosters leadership development among program directors. Its approach is heavily organized around performance indicators and feedback mechanisms embedded in directors' and supervisors' workflow. This is an example of integrating measurement and supports to improve identified professional competencies within systems of implementation and workflow management—an approach that is rare in educators' PD and training.

### ► Provide necessary supports for the PD workforce

PD's success depends in large part on the people who train and coach teachers. This means hiring, training, and supporting the PD workforce. But little research has examined these elements of program delivery, and many evidence-based PD models fail to provide much detail about them. Among evidence-based PD models that do provide such detail, this workforce typically consists of experienced ECE teachers, often with master's degrees, who have relatively extensive training and ongoing support in the particular PD model (McCollum, Hemmeter, & Hsieh,

2011; Piasta et al., 2012; Powell, Diamond, Burchinal, & Koehler, 2010). Lloyd & Modlin (2012), reporting on how they delivered three coaching models in Head Start programs, suggest that successful coaches have three major attributes: knowledge of the coaching model, general coaching and consultation skills, and knowledge of early childhood development and teaching. The Ounce of Prevention leadership initiative's job-embedded training systems, described above, constitute an example of how systems (technology, measurement, management) can also support PD providers.

In most cases, evidence-based models include fairly intensive initial training as well as weekly supervision of coaches (Isner et al., 2011). This is rarely the case in the field. For example, within the scope of Head Start's large-scale initiative on mentor-coaching, most Head Start coaches report having had some training and supervision, but very little of it was specific to coaching (Howard et al., 2013). Only 16% of the coaches described receiving any specific training related to coaching. By contrast, in the MTP evaluations, coaches participated in a weeklong training session focused on CLASS, the MTP coaching model, and use of the MTP website to support teachers; all coaches became reliable on the CLASS instrument. Coaches received ongoing help from dedicated coach-support staff, including booster training, weekly phone calls to individual coaches, and group coaching calls. Group and individual calls every two weeks give coaches a forum for sharing successes and challenges of the job.

Coaching, particularly when it follows standardized and structured models, can be highly effective for improving teachers' practices in the classroom, even in larger-scale implementations (Bierman et al., 2008; Cunningham, Zibulsky, & Callahan, 2009; Dickinson & Caswell, 2007). But coaching requires sufficient attention to supervision, adherence to standardized protocols, and use of a model that makes teachers and coaches feel effective and motivated to participate. Yet Isner et al. (2011), in their study of coaching as a part of QRIS, report that very few programs used any formal manual or set of materials to guide coaches' daily practice.

### ► **Harness higher education as a workforce development and PD delivery system that delivers results**

Despite the potential for coursework or degrees in higher education to improve teacher impacts, there is no consistently identifiable link between the two. And yet, as we describe above, there are numerous examples of courses that have led to improvements in practice. What supports are needed so that these exemplars of impact and success can be used at greater scale?

As one example, a series of follow-up investigations related to the course based on CLASS examined the supports needed to deliver the course in 15 sections, with sufficient fidelity to support impacts on teacher practice (LoCasale-Crouch et al., 2011). The list was long. Two course coordinators provided training and implementation support to 14 instructors. Course instructors were trained to achieve reliability on CLASS and on course content and implementation, to ensure consistent delivery. Before teaching each unit, instructors and course coordinators

met online to review upcoming activities, including PowerPoint slides, the instructor's manual, readings, in-class activities, homework assignments, and exams. Instructors completed a written assignment related to each unit, showing evidence that they understood and were comfortable with the material. Course coordinators held weekly individual support calls and periodic group calls with course instructors that were focused on clarifying content, implementation, and sharing successes and challenges in teaching the course. On five occasions, course instructors videotaped themselves teaching the planned lesson and received written feedback that was discussed in detail during the weekly call. As the course went on, the instructors improved and became more consistent in observed implementation.

Although the amount of support was considerable, it should also be noted that these supports were highly targeted and delivered using distal means across 15 sections at 10 different institutions. Under these conditions, 14 instructors delivered a common course with high degrees of skill, fidelity, and implementation quality, all leading to significant impact on teachers' practices in the classroom (Downer et al., 2011; Hamre et al., in press; LoCasale-Crouch et al., 2011). Embedded in a system of appropriate focus, structure, and support, course-based PD can be implemented with high fidelity at scale.

### ► **Use data to target and improve PD**

Although some programs collect child-outcome data and use it to support individualized approaches to instruction, fewer of them use these data at the program level to drive PD. Programs tend to lack refined indicators of teacher knowledge or competencies to use such data to tailor workforce development initiatives to individuals' profiles of knowledge and skill. Programs also often struggle to ask the right questions of their data, whether related to child outcomes or the workforce, and they often lack expertise in the technical skills required to efficiently collect, maintain, analyze, and interpret data (Crawford, Tucker, Van Horne, & Landry, 2016).

However, data can not only help to focus PD but can also track its implementation and success. Lloyd & Modlin (2012) describe a simple but effective method for supporting the coaching delivered as a part of the Head Start CARES project. They use brief online surveys, logs, and fidelity reports to help support technical assistance and management in their monitoring of coaching implementation. Similar systems are provided with the scaled-up version of MTP (Early et al., 2017). Even the simplest information, such as logs of the frequency of contacts between teachers and coaches, can be powerful ways to improve the intensity of coaching if they are used to monitor coaches' efforts and provide feedback. To the extent that PD is delivered online, the web interface and backend can provide useful data for enabling strong implementation supports for teachers, course instructors, and coaches (LoCasale et al., 2016). As states build systems of PD support online and link them to various forms of credentialing (including micro-credentialing), the result can be more fully integrated alignment of teachers' PD needs and goals, PD inputs to teachers, supports for effective delivery (by coaches, instructors, or web systems), and structures that codify and encourage teachers' participation and progress.

### ► Link workforce development systems and incentive structures

Most states, school districts, and Head Start programs require only that teachers complete a certain number of clock hours of PD each year, ranging from over 100 to 15 or fewer (Barnett et al., 2016). All states give programs flexibility in how these hours are allocated, reducing the likelihood that those hours (or any effective PD approach) will drive program improvements. One way states have tried to tighten the link between PD hours and impact is to require teachers, directors, and/or coaches to articulate clear PD plans and then evaluate those plans (Rous, Grove, Cox, Townley, & Crumpton, 2008). State workforce registry systems are typically limited to tracking members of the ECE workforce (often volunteer participants), their credentials, and the PD they have attended (Ryan & Whitebook, 2012). However, registry systems are being developed that codify individual teachers' records of acquired PD (National Registry Alliance, 2013a) and perhaps even the competencies they attain, which will mean greater capability to identify and encourage effective PD as well to tie those experiences to accrued competence and certifications.

### ► Certify PD providers

The skills and impact of those who provide PD support to teachers and programs vary widely (Soliday-Hong, Walters, & Mintz, 2011), and there are very few systems for documenting their expertise and effectiveness. Although almost half of the states have developed tracking systems for PD providers (Institute of Medicine and National Research Council, 2015), none have effectiveness metrics or standard certifications and training. Some have moved beyond tracking to comprehensive training and certification requirements for providers. For example, anyone who receives funding from the state of Pennsylvania to offer training has to participate in the Pennsylvania Quality Assurance System, which includes online coursework and a review of professional development activities (Hong et al.).

In some states, PD providers must register and complete training (National Registry Alliance, 2013b), but these systems are typically voluntary and their requirements are not particularly stringent. Clearly, PD providers and coaches need more intensive training and certification programs. Examples on which to build include the University of Colorado Early Childhood Coaching Certificate program, a three-course series that focuses on developing specific coaching and organizational change skills. Yet, despite some promising developments, such programs are the exception; PD staff hired by preschool programs rarely have robust and ongoing training.

## CONCLUSION

---

**If we wish to narrow intransigent gaps in children’s experiences and outcomes, research points to a clear need for systems of program design, implementation, and improvement that span the period from birth through preschool and up to third grade.**

We cannot improve quality and impact in the U.S. early education and care sector simply through renewed appreciation for workforce development. Rather, if we wish to narrow intransigent gaps in children’s experiences and outcomes, research points to a clear need for systems of program design, implementation, and improvement that span the period from birth through preschool and up to third

grade. These systems must not only select and disseminate proven-effective models of professional development, they also must meet the conditions, such as incentives, data, and certification regimes, that allow PD models to be scaled with fidelity. With increased use of technology to deliver PD online as well as continuing refinement of PD models to deliver relevant knowledge and training of practice-focused skills, a future of individualized PD pathways, stackable credentials, state registries, and even increased compensation may not be far off.

## References

---

- Ansari, A., & Pianta, R. (2018). Classroom age composition and the early learning of preschoolers. *Journal of Educational Research*, 112(2), 234-242. doi: 10.1080/00220671.2018. 1514356
- Barnett, W. S., Friedman-Krauss, A. H., Gomez, R. E., Horowitz, M., Weisenfeld, G. G., Clarke Brown, K., & Squires, J. H. (2016). *The State of Preschool 2015: State Preschool Yearbook*. New Brunswick, NJ: National Institute for Early Education Research.
- Baroody, A. J., Clements, D. H., & Sarama, J. (2019). Teaching and learning mathematics in early childhood programs. In C. Brown, M. B. McMullen, & N. File (Eds.), *Handbook of Early Childhood Care and Education* (1st ed., pp. 329-353). Hoboken, NJ: Wiley Blackwell Publishing.
- Barton, E. E., Fuller, E. A., & Schnitz, A. (2016). The use of email to coach preservice early childhood teachers. *Topics in Early Childhood Special Education*, 36(2), 78-90. doi: 10.1177/0271121415612728
- Berkel, C., Mauricio, A., Schoenfelder, E., & Sandler, I.N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science*, 12(1), 23-33. doi: 10.1007/s11121-010-0186-1
- Bierman, K.L., Nix, R.L., Greenberg, M.T., Blair, C., & Domitrovich, C. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in Head Start REDI program. *Development and Psychopathology*, 20, 821-843. doi: 10.1017/S0954579408000394
- Birman, B.F., Desimone, L., Porter, A.C., & Garet, M.S. (2000). Designing professional development that works. *Educational Leadership*, 57(8), 28-33.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3-15. doi: 10.3102/0013189X033008003
- Chen, J.Q., & McCray, J. (2012). The what, how, and why of effective teacher professional development in early mathematics education. *NHSA Dialog*, 15(1), 113-121.
- Cochran-Smith, M., & Zeichner, K.M. (2005). *Studying teacher education: The report of the AERA panel on research and teacher education*. Washington, DC: AERA.
- Connors, M.C., Pacchiano, D.M., Manos, E., & Horsley, H.L. (2018, March). *Building the infrastructure: Leveraging research partnerships to improve and scale an instructional leadership professional development program*. Paper presented at Society for Research on Educational Effectiveness Spring 2018 Conference, Washington, DC.
- Crawford, A., Zucker, T., Van Horne, B., & Landry, S. (2016). Integrating professional development content and formative assessment with the coaching process: The Texas School Ready model. *Theory into Practice*, 56(1), 55-65.
- Cunningham, A.E., Zibulsky, J., & Callahan, M. (2009). Starting small: Building preschool teacher knowledge that supports early literacy development. *Reading and Writing*, 22(4), 487-510. doi: 10.1007/s11145-009-9164-z
- 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. doi: 10.1016/S0272-7358(97)00043-3
- Desimone, L. M., & Garet, M. S. (2015). Best practices in teachers' professional development in the United States. *Psychology, Society and Education*, 7(3), 252-263.
- Dickinson, D., & Caswell, L. (2007). Building support for language and early literacy in preschool classrooms through in-service professional development: Effects of the Literacy Environment Enrichment Program (LEEP). *Early Childhood Research Quarterly*, 22, 243-260. doi: 10.1016/j.ecresq.2007.03.001

## CHAPTER 5 IMPROVING QUALITY AND IMPACT THROUGH WORKFORCE DEVELOPMENT AND IMPLEMENTATION SYSTEMS

- Domitrovich, C.E., Gest, S.D., Gill, S., Bierman, K.L., Welsh, J.A., & Jones, D. (2009). Fostering high quality teaching with an enriched curriculum and professional development support: The Head Start REDI program. *American Educational Research Journal*, 46(2), 567-597. doi: 10.3102/0002831208328089
- Donohue, C., Fox, S., & Torrence, D. (2007). Early childhood educators as elearners: Engaging approaches to teaching and learning online. *Young Children*, 34-40.
- Downer, J.T., Pianta, R.C., Fan, X., Hamre, B.K., Mashburn, A., & Justice, L. (2011). Effects of web-mediated teacher professional development on the language and literacy skills of children enrolled in prekindergarten programs. *NHSA Dialog: Research-to-Practice Journal for the Early Childhood Field*, 14(4), 189-212.
- Downer, J.T., Pianta, R.C., Burchinal, M., Field, S., Hamre, B.K., LoCasale-Crouch, J., ... Scott-Little, C. (2011). *Coaching and coursework focused on teacher-child interactions during language/literacy instruction: Effects on teacher beliefs, knowledge, skills, and practice*. Manuscript in preparation.
- Early, D.M., Maxwell, K.L., Ponder, B.D., & Pan, Y. (2017). Improving teacher-child interactions: A randomized control trial of Making the Most of Classroom Interactions and My Teaching Partner professional development models. *Early Childhood Research Quarterly*, 38, 57-70. doi: 10.1016/j.ecresq.2016.08.005
- Early, D.M., Maxwell, K.L., Burchinal, M., Alva, S., Bender, R.H., Bryant, D., ... Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development*, 78(2), 558-580. doi: 10.1111/j.1467-8624.2007.01014.x
- Fox, L., Hemmeter, M.L., & Snyder, P.S. (2014). *Teaching Pyramid Observation Tool for Preschool Classrooms (TPOT), Research Edition*. Baltimore, MD: Paul H. Brookes.
- Fukkink, R.G., & Lont, A. (2007). Does training matter? A meta-analysis and review of caregiver training studies. *Early Childhood Research Quarterly*, 22, 294-311. doi: 10.1016/j.ecresq.2007.04.005
- Garet, M.S., Porter, A.C., Desimone, L., Birman, B.F., & Suk Yoon, K. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915-945.
- Gerde, H.K., Duke, N.K., Moses, A.M., Spybrook, J., & Shedd, M.K. (2014). How much for whom? Lessons from an efficacy study of modest professional development for child care providers. *Early Education and Development*, 25, 421-441. doi: 10.1080/10409289.2013.788424
- Hamre, B.K., Pianta, R.C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J.T., ... & Scott-Little, C. (2012). A course on effective teacher-child interactions: Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal*, 49(1), 88-123. doi: 10.3102/0002831211434596
- Hamre, B. K., Pianta, R. C., Mashburn, A. J., & Downer, J. T. (2012). Promoting young children's social competence through the preschool PATHS curriculum and MyTeachingPartner professional development resources. *Early Education and Development*, 23(6), 809-832.
- Hamre, B.K., Downer, J.T., Jamil, F.M., & Pianta, R. (2012). Enhancing teachers' intentional use of effective interactions with children: Designing and testing professional development interventions. In R. Pianta (Ed.), *Handbook of early childhood education* (507-532). New York, NY: Guilford Press.
- Hamre, B.K., Pianta, R. C., Burchinal, M., DeCoster, J., & Downer, J. (2015, March). *Variation in impacts of MyTeachingPartner coaching model on preschool children's development*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Philadelphia, PA.
- Hamre, B.K., Pianta, R.C., Burchinal, M., & Downer, J.T. (in press). A course on supporting early language and literacy development through effective teacher-child interactions: Effects on teacher beliefs, knowledge and practice. *American Educational Research Journal*.

Hemmeter, M.L., Fox, L., & Snyder, P. (2013). A tiered model for promoting social-emotional competence and addressing challenging behavior. In V. Buysse & E. Peisner-Feinberg (Eds.), *Handbook of response to intervention in early childhood* (85-101). Baltimore, MD: Paul H. Brookes.

Hemmeter, M.L., Hardy, J.K., Schnitz, A.G., Adams, J.M., & Kinder, K.A. (2015). Effects of training and coaching with performance feedback on teachers' use of pyramid model practices. *Topics in Early Childhood Special Education, 35*(3), 144-156. doi: 10.1177/0271121415594924

Howard, E.C., Rankin, V.E., Fishman, M., Hawkinson, L.E., McGroder, S.M., Helsel, F.K., et al. (2013). *The Descriptive Study of the Head Start Early Learning Mentor Coach Initiative*. OPRE Report #2014-5a; Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Howes, C., Galinsky, E., & Kontos, S. (1998). Child care caregiver sensitivity and attachment. *Social Development, 7*(1), 25-36. doi: 10.1111/1467-9507.00048

Institute of Medicine (IOM) and National Research Council (NRC). 2015. *Transforming the workforce for children birth through age 8: A unifying foundation*. Washington, DC: National Academies Press.

Isner, T., Tout, K., Zaslow, M., Soli, M., Quinn, K., Rothenberg, L., & Burkhauser, M. (2011). *Coaching in early care and education programs and Quality Rating and Improvement Systems (QRIS): Identifying promising features*. Report produced for the Children's Services Council of Palm Beach County. Washington, DC: Child Trends.

Kontos, S., Howes, C., & Galinsky, E. (1996). Does training make a difference to quality in family child care? *Early Childhood Research Quarterly, 11*(4), 427-445. doi: 10.1016/S0885-2006(96)90016-2

Kraft, M.A., Blazar, D., & Hogan, D. (2016). The effect of teaching coaching on instruction and achievement: A meta-analysis of the causal evidence. Working paper, Brown University, Providence, RI.

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. doi 10.1037/a0016746

Landry, S.H., Crawford, A., Gunnewig, S., & Swank, P.R. (2002). *Teacher Behavior Rating Scale*. Unpublished research instrument, University of Texas Health Science Center at Houston, Center for Improving the Readiness of Children for Learning and Education.

Landry, S.H., Swank, P.R., Anthony, J.L., & Assel, M.A. (2011). An experimental study evaluating professional develop activities within a state funded pre-kindergarten program. *Reading and Writing, 24*(8), 971-1010. doi: 10.1007/s11145-010-9243-1

Lloyd, C.M., & Modlin, E.L. (2012). *Coaching as a Key Component in Teachers' Professional Development: Improving Classroom Practices in Head Start Settings*. OPRE Report 2012-4, Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

LoCasale-Crouch, J., Hamre, B., Roberts, A., & Neesen, K. (2016). If You Build It, Will They Come? Predictors of Teachers' Participation in and Satisfaction with the Effective Classroom Interactions Online Courses. *The International Review of Research in Open and Distributed Learning, 17*(1).

LoCasale-Crouch, J., Kraft-Sayre, M., Pianta, R., Hamre, B., Downer, J., Leach, A., ... Scott-Little, C. (2011). Implementing an early childhood professional development course across 10 sites and 15 sections: Lessons learned. *NHSA Dialog, 14*(4), 275-292.

Markussen-Brown, J., Juhl, C.B., Piasta, S.B., Bleses, D., Hojen, A., & Justice, L.M. (2017). The effects of language- and literacy-focused professional development on early educators and children: A best-evidence meta-analysis. *Early Childhood Research Quarterly, 38*, 97-115. doi: 10.1016/j.ecresq.2016.07.002

## CHAPTER 5 IMPROVING QUALITY AND IMPACT THROUGH WORKFORCE DEVELOPMENT AND IMPLEMENTATION SYSTEMS

- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., ... & Howes, C. (2008). Measures of classroom quality in pre-kindergarten and children's development of academic, language and social skills. *Child Development, 79*(3), 732-749. doi: 10.1111/j.1467-8624.2008.01154.x
- Mashburn, A.J., Downer, J.T., Hamre, B.K., Justice, L.M., & Pianta, R.C. (2010). Consultation for teachers and children's language and literacy development during pre-kindergarten. *Applied Developmental Science, 14*, 179–196. doi: 10.1080/10888691.2010.516187
- McCullum, J.A., Hemmeter, M.L., & Hsieh, W-Y. (2011). Coaching teachers for emergent literacy instruction using performance-based feedback. *Topics in Early Childhood Special Education, 33*(1), 28-37. doi: 10.1177/0271121411431003
- McCormick Center for Early Childhood Leadership. (2016). *Quality standards drive professional development opportunities*. Wheeling, IL: National-Louis University, McCormick Center for Early Childhood Leadership. Retrieved from <http://mccormickcenter.nl.edu/wp-content/uploads/2016/11/RN-Fall-2016-final.pdf>
- Mendive, S., Weiland, C., Yoshikawa, H., & Snow, C. (2016). Opening the black box: Intervention fidelity in a randomized trial of a preschool teacher professional development program. *Journal of Educational Psychology, 108*(1), 130-145. doi: 10.1037/edu0000047
- National Child Care Information and Technical Assistance Center (NCCITA) and the National Association for Regulatory Administration (NARA). (2010). *The 2008 child care licensing study*. Lexington, KY: National Association for Regulatory Administration.
- National Registry Alliance. (2013a). *State of registries survey 2012: A survey of the nation's early childhood and school-age registries*. Washington, DC: National Registry Alliance.
- National Registry Alliance. (2013b). *2012 state of the nation's training and trainer registries and approval systems*. Washington, DC: National Registry Alliance.
- Neuman, S.B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal, 46*(2), 532-566. doi: 10.3102/0002831208328088
- Pianta, R. C., DeCoster, J., Cabell, S.Q., Burchinal, M., Hamre, B.K., Downer, J.T., & Howes, C. (2014). Dose-response relations between preschool teachers' exposure to components of professional development and increases in quality of their interactions with children. *Early Childhood Research Quarterly, 29*(4), 499-508. doi: 10.1016/j.ecresq.2014.06.001
- Pianta, R.C., Hamre, B.K., Downer, J.T., Burchinal, M., Williford, A., LoCasale-Crouch, J., ... & Scott-Little, C. (2017). Early childhood professional development: Coaching and coursework effects on indicators of children's school readiness. *Early Education and Development, 28*(8). doi: 10.1080/10409289.2017.1319783
- Pianta, R.C., Hamre, B.K., & Hadden, S. (2012) Scaling-up effective professional development. In C. Howes, B. Hamre & R. Pianta (Eds.), *Effective early childhood professional development: Improving teacher practice and child outcomes*. Baltimore, MD: Paul H. Brookes.
- Pianta, R. C., & Hofkens, T. (2018). The Bush-Obama agenda for education research and innovation: Major investment, mixed returns. In F. M. Hess, & M. Q. McShane (Eds.), *Bush-Obama school reform* (87-106). Cambridge, MA: Harvard Education Press.
- Pianta, R.C., La Paro, K.M., & Hamre, B. (2008). *The classroom assessment scoring system, Pre-K version*. Baltimore, MD: Paul H. Brookes.
- Piasta, S.B., Justice, L.M., Cabell, S.Q., Wiggins, A.K., Turnbull, K.P., & Curenton, S.M. (2012). Impact of professional development on preschool teachers' conversational responsivity and children's linguistic productivity and complexity. *Early Childhood Research Quarterly, 27*, 387-400. doi: 10.1016/j.ecresq.2012.01.001

- Piasta, S.B., Logan, J.A., Pelatti, C.Y., Capps, J.L., & Petrill, S.A. (2015). Professional development for early childhood educators: Efforts to improve math and science learning opportunities in early childhood classrooms. *Journal of educational psychology*, 107(2), 407-422.
- Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M. J. (2010). Effects of an early literacy professional development intervention on head start teachers and children. *Journal of Educational Psychology*, 102(2), 299-312. doi: 10.1037/a0017763
- Raver, C.C., Jones, A.S., Li-Grining, C.P., Metzger, M., Smallwood, K., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in head start settings. *Early Childhood Research Quarterly*, 23(1), 10-26. doi: 10.1016/j.ecresq.2007.09.001
- Resnicow, K., Davis, M., Smith, M., Lazarus-Yaroch, A., Baranowski, T., Baranowski, J., ... Wang, T. (1998). How best to measure implementation of school health curricula: A comparison of three measures. *Health Education Research*, 13(2), 239-250. doi: 10.1093/her/13.2.239
- Rous, B., Grove, J., Cox, M., Townley, K., & Crumpton, G. (2008). *The impact of the Kentucky professional development framework on child care, head start and public preschool classroom quality and child outcomes*. Lexington, KY: University of Kentucky, Human Development Institute.
- Ryan, S., & Whitebook, M. (2012). More than teachers: The early care and education workforce. In R.C. Pianta (Ed.), *Handbook of early childhood education* (92-110). New York, NY: Guilford Press.
- Schachter, R.E. (2015). An analytic study of the professional development research in early childhood education. *Early Education and Development*, 26(8), 1057-1085.
- Scott-Little, C., La Paro, K. M., Thomason, A. C., Pianta, R. C., Hamre, B., Downer, J., . . . Howes, C. (2011). Implementation of a course focused on language and literacy within teacher-child interactions: Instructor and student perspectives across three institutions of higher education. *Journal of Early Childhood Teacher Education*, 32(3), 200-224.
- Snyder, P., Hemmeter, M.L., Artman Meeker, K., Kinder, K., Pasia, C., & McLaughlin, T. (2012). Characterizing key features of the early childhood professional development literature. *Infants and Young Children*, 25(3), 188-212. doi: 10.1097/IYC.0b013e31825a1ebf
- Soliday-Hong, S. L., Walters, T., & Mintz, T. (2011). Professional development systems for early childhood educators within a state and federal policy context. In C. Howes, & R. C. Pianta (Eds.), *Foundations for teaching excellence: Connecting early childhood quality rating, professional development, and competency systems in states* (117-132). Baltimore, MD: Paul H. Brookes.
- Taggart, V.S., Bush, P.J., Zuckerman, A.E., & Theiss, P.K. (1990). A process evaluation of the District of Columbia "Know Your Body" project. *Journal of School Health*, 60(2), 60-66. doi: 10.1111/j.1746-1561.1990.tb05907.x
- Tout, K., Halle, T.G., Datta, A.R., & Snow, K. (2015, June). *A National Portrait of Quality in Centers: Recent Findings from the National Survey of Early Care and Education*. NAEYC Professional Development Institute, New Orleans, LA.
- U.S. Department of Education and U.S. Department of Health and Human Services. (2014). *The Race to the Top - Early Learning Challenge Year Two Progress Report*. Washington, DC.
- Werner, C.D., Linting, M., Vermeer, H. J., & Van IJzendoorn, M.H. (2016). Do intervention programs in child care promote the quality of caregiver-child Interactions? A meta-analysis of randomized controlled trials. *Prevention Science*, 17(2), 259-273.
- Whalen, S., Horsley, H., Parkinson, K., & Pacchiano, D. (2016). A development evaluation study of a professional development initiative to strengthen organizational conditions in early education settings. *Journal of Applied Research on Children: Informing Policy for Children at Risk*, 7(2), Article 9.
- Whitebook, M., Bellm, D., Lee, Y., & Sakai, L. (2005). *Time to revamp and expand: Early childhood teacher preparation programs in California's institutions of higher education*. Berkeley, CA: Center for the Study of Child Care Employment.

## CHAPTER 5 IMPROVING QUALITY AND IMPACT THROUGH WORKFORCE DEVELOPMENT AND IMPLEMENTATION SYSTEMS

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, Berkeley.

Williford, A.P., Carter, L.M., Maier, M.F., Hamre, B.K., Cash, A., Pianta, R.C., & Downer, J.T. (2017). Teacher engagement in core components of an effective, early childhood professional development course: Links to changes in teacher-child interactions. *Journal of Early Childhood Teacher Education*, 38(1), 102-118. doi: 10.1080/10901027.2016.1269028

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M., Espinosa, L., Gormley, W., ... & Zaslow, M. (2013). *Investing in our future: The evidence base on preschool education*. Washington, DC: Society for Research in Child Development and the Foundation for Child Development.

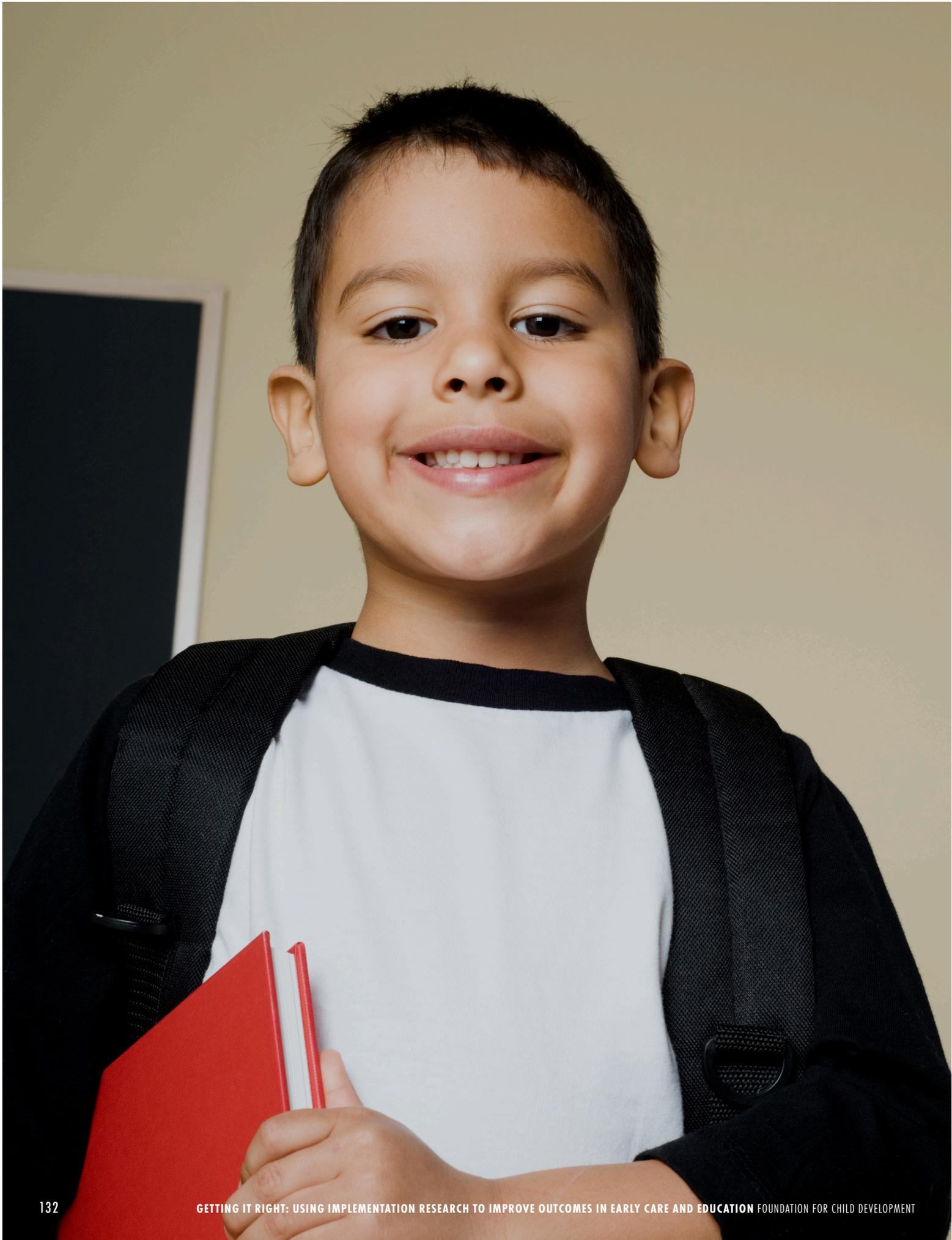
Zaslow, M., Tout, K., Halle, T., Whittaker, J. V., & Lavelle, B. (2010). *Toward the identification of features of effective professional development for early childhood educators: Literature review*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Social Development.

**SECTION 2, CHAPTER 6**

**ADDRESSING EQUITY IN THE  
ECE CLASSROOM: EQUAL ACCESS  
AND HIGH QUALITY FOR  
DUAL LANGUAGE LEARNERS**

---

Linda M. Espinosa, Ph.D., University of Missouri-Columbia



The early childhood education (ECE) profession has a long-standing commitment to the principle of equity and to antidiscriminatory practices, as the recent position statement from the National Association for the Education of Young Children (2019) makes clear. This position statement, combined with supporting curriculum and assessment materials, promotes equal access to high-quality early education and affirms the value of all children and families for their unique talents and cultural and linguistic strengths. One particular group of diverse children, dual language learners (DLLs)—meaning children who are aged 0-5 and speak a language other than English in the home—face a number of challenges that contribute to decreased educational attainment and have implications for ECE educational equity (Castro, Espinosa, & Páez, 2011).

Research shows that all young children can learn more than one language during the ECE years and that doing so carries significant linguistic, academic, social, and cognitive advantages (NASEM, 2017). Yet many dual language learners evidence achievement gaps in comparison to native English speakers (EOs), suggesting that ECE educators need to adopt new strategies for actualizing the academic and intellectual potential of DLLs. To design effective educational approaches for DLLs, we must first understand what typical development and school readiness looks like for these children, what factors contribute to their growth and learning, and what teaching practices and classroom conditions best support their achievement. In this chapter, I propose that we shift the conceptual frame for understanding and improving instructional practices for DLLs by essentially broadening critical pedagogical knowledge and how to apply it.

To provide equitable early education to linguistically diverse children, ECE teachers must consistently implement a set of instructional adaptations across multiple settings. One core necessity here is to recognize that these children are learning content or conceptual knowledge at the same time that they are also learning the language in which that content or concept is expressed. Thus instructional approaches that focus on monolingual English speakers need to be adapted and enhanced (Castro, Espinosa, & Páez, 2011; NASEM, 2017) to build on what children already know in their first language while they are also adding English. This chapter outlines the research on the benefits of early bilingualism and presents specific strategies that all ECE teachers can implement that will support DLLs' acquisition of English while also maintaining their home language. I first summarize the research on early bilingualism and then outline instructional adaptations based on current scientific evidence on how to support improved outcomes for DLLs.

**Instructional approaches that focus on monolingual English speakers need to be adapted and enhanced to build on what children already know in their first language while they are also adding English.**

Why do we need high-quality ECE for DLLs? One of the driving forces behind publicly funded ECE programs, such as Head Start and state prekindergarten programs, has been compensatory education. These programs have been largely designed to provide early learning experiences that promote “school readiness” for children from low-income homes, many of whom are minorities and do not speak English in the home. In fact, Head Start’s stated mission is to promote “the school readiness of young children from low-income families by enhancing their cognitive, social, and emotional development” (Office of Head Start, 2015). Almost 30% of U.S. 4-year olds are served by state prekindergarten programs (Barnett, Carolan, Squires, Clarke Brown, & Horowitz, 2015), most of which have income eligibility requirements and are focused on increasing vulnerable preschoolers’ access to high-quality ECE. Based on both historical and current empirical research that has demonstrated that children who attend a year or two of high-quality ECE have better oral language, literacy, and mathematics scores at kindergarten entry than their peers who do not have such experiences (Yoshikawa et al., 2015), government at all levels has been seeking to expand access to and improve the quality of early education (National Institute for Early Education Research, 2017). These efforts have primarily targeted children from low-income families, with the ultimate goal of reducing the achievement gap at kindergarten entry and improving long-term school success.

Researchers have stressed that high quality is important to achieve improved academic skills that are both discernable at the end of prekindergarten and sustained into the elementary school years. As Dale Farran notes in this volume, a central element of high-quality education during the early years is frequent, warm, responsive, engaging interactions between adults and children that include multiple turn-taking. Ensuring these kinds of interactions for children who are not native English speakers and whose English language skills are not well developed is difficult in ECE settings. Researchers and practitioners are asking a range of questions to address these challenges. What language should be used during these individual and group interactions? At what age should young children be exposed to a second language? How much of each language should be used throughout the preschool day? What qualifications should ECE staff have to best meet the needs of children who understand very little English? What strategies can monolingual English-speaking teachers use when working with children who do not speak or understand English? Do state and local learning standards apply equitably to all language speakers? What are reasonable expectations for language growth? And, finally, how can ECE staff assess progress when children have limited English skills?

The growth of DLLs in the child population has meant that many ECE settings, such as Head Start and state prekindergarten programs, now serve large numbers of families and children who primarily speak languages other than English. Demographics demonstrate the increasing linguistic diversity of our children and families. Although many states do not collect data on their preschool DLLs (National Institute for Early Education Research, 2018), the U.S. Census Bureau estimates that nearly one-third of all children ages birth to 8 are growing up with exposure to more than one language in the home (Park, Zong, & Batalova, 2018). The Office of Head Start (2017) reports that more than 30% of preschool children in their programs are considered DLLs, and in the state of California, 60% of

children ages 0 to 5 are so identified (First Five California, 2017). More than 130 different languages have been identified in the Head Start child population; more than 80% of all Head Start classrooms serve DLLs, who in many cases speak multiple languages. Unfortunately, ECE teachers who speak more than one language remain in short supply, making up only about 15% of the workforce (Park, McHugh, Zong, & Batalova, 2015).

The substantial and persistent achievement gap between DLLs and native English speakers is of concern to researchers, educators, and policymakers across the U.S. In many studies, DLLs show language gaps during infancy, although language is almost always assessed only in English in these studies and DLLs have had fewer opportunities to learn English (Fuller, Bein, Kim, & Rabe-Hesketh, 2015). They perform significantly below their English-only peers at kindergarten entry and have much lower reading and math scores at third grade. Many are classified as long-term English learners (ELs) during upper grades and have little access to the general curriculum and a higher probability of dropping out of school (NASEM, 2017; Olsen, 2010).

To effectively provide educational equity and high-quality ECE for DLLs, we must define and put into practice effective program language models, specific instructional practices that scaffold language interactions for DLLs, instruments and methods for ongoing assessment, and ECE teacher qualifications. Fortunately, scientific knowledge about how a young child learns a second language and what constitutes best practice in ECE for DLLs has expanded greatly during the past decade (NASEM, 2017). Yet many questions about specific approaches and instructional practices remain.

## REJECTING THE DEFICIT APPROACH TO DUAL LANGUAGE LEARNING

---

Historically, most research examining the growth, progress, and achievement of DLLs has focused on differences between DLLs and non-DLLs, judging DLLs' performance using norms designed for English-only populations without considerations for the unique linguistic and developmental trajectories of children whose first language is not English (Center for Early Care and Education Research—Dual Language Learners, 2011). This approach has often led to a “deficit perspective” that views DLLs as having less potential and fewer academic abilities than their monolingual English peers because of their lack of English proficiency. In fact, policymakers have sometimes referred to “the extra burden” of having to learn two languages during the early years. The deficit perspective, however, often negatively affects teachers' views of DLLs' potential, and it is, moreover, contradicted by current research.

The scientific consensus is that children who become fully proficient in both their home language and English are likely to reap benefits in cognitive, social, academic, and professional outcomes and to be protected from brain decline at older ages (NASEM, 2017). This suggests we should view the development of DLLs through the powerful advantages of having more than one language. The assets associated with bilingualism and biliteracy have been well documented and should be recognized and supported.

**The scientific consensus is that children who become fully proficient in both their home language and English are likely to reap benefits in cognitive, social, academic, and professional outcomes and to be protected from brain decline at older ages.**

All who work with children who speak a language other than English in the home must recognize that DLLs' development differs in significant ways from that of their native English-speaking peers due to the unique context and societal circumstances of their upbringing. For example, although more than 90% of DLLs are born in the United States (NASEM, 2017), often one or both of their parents were born elsewhere. Many of these families have recently immigrated to the United States and may be unfamiliar with social and cultural norms or school expectations. Some of them have experienced trauma associated with migrating to the United States, which can have negative cognitive and social consequences for child development (Yoshikawa, 2011). And by definition, the families of

DLLs speak a language other than English in the home, a characteristic that can lead to social isolation and, in some cases, can create mixed feelings or even a sense of shame for the children (Halgunseth, Jia, & Barbarin, 2013). Culture-specific parenting goals, values, and practices that vary across ethnic groups can contribute to inaccurate perceptions of DLLs' early social, language, and literacy potential. For instance, among Latino families, culturally specific parenting concepts such as *familismo* (family), *respeto* (respect), and being *bien educado* (well educated) (Halgunseth et al., 2013) emphasize the importance of harmonious relationships with others, respect for adult authority, prioritizing the needs of the family, and conducting oneself in a manner that does not bring shame on the family or community. Other values that children are exposed to early in life may include a focus on group or collective well-being rather than individualism, individualism being an attribute stressing independence and self-reliance that is commonly emphasized in American schools (Small, 2002). These contrasting early socialization practices can lead to patterns of behavior that are inconsistent with ECE program goals, such as being reluctant to stand out as the only child who knows the answer, and they can give teachers' a misleading impression of DLLs knowledge level.

Family members' beliefs about exposure to English and continued use of the home language also affect their children's language learning and academic success (Billings, 2009). Some may view the home language as critical for maintaining ties to the family's cultural heritage and connections with family members in their countries of origin. Conversely, newly arrived immigrant families may prize the rapid acquisition of English over maintenance of their heritage language and encourage their children to speak only English. Thus, beliefs and goals about cultural and language maintenance can play a key role in how much exposure and opportunity children have to use their two languages.

The family contexts and early learning environments of DLLs vary widely, and thus they should not be considered a homogeneous group or only in comparison to their English-only peers. Sociocultural and demographic variables such as language spoken in the home, age at first exposure to English, family socioeconomic status, and country of origin can all influence children's proficiency and early literacy skills in both the home language and English

(NASEM, 2017). To understand each DLL's language status and educational needs, ECE teachers need in-depth knowledge of their family circumstances, values, and culture. Specifically, ECE personnel must expand their thinking beyond simple comparisons between DLLs and English-only children and not use norms or learning trajectories based on English-only learners. All ECE program leaders need to design tools and methods to collect important information about DLLs' background (e.g., the age of acquisition of each language, the extent and nature of exposure to each language, and key family characteristics) as well as family histories that go beyond the typical home language survey.

Finally, the amount and quality of DLLs' exposure to and usage of their two languages are also important features of early development that impact later school success. Multiple studies have shown that preschoolers' and school-age children's exposure to the home language supports their development of that language (Hammer et al., 2012). Use of the child's first language in the home or in school does not appear to affect the rate or level of English acquisition. However, emphasizing English in the ECE setting does appear to undermine DLLs' continued development of the home language. This is likely due to the higher value given to English proficiency at school and in the broader social context. Given research findings about the impact of exposure to their two languages at home and in school, we should devote attention to the amount and quality of exposure DLLs experience in each language.

What follows is a discussion of some recent findings and conclusions about dual language development during the early years and specific classroom practices that have empirical evidence of efficacy for linguistically diverse children. Hopefully, if we clearly and explicitly communicate how young children acquire and benefit from exposure to more than one language and describe in detail which practices have shown pedagogical promise, we can produce more equitable and higher-quality ECE for DLLs.

## CURRENT RESEARCH ON EARLY BILINGUAL DEVELOPMENT

---

As knowledge concerning DLLs' language development has grown, it has increasingly been used as a foundation to support and guide ECE practice. Several strands of research from multiple disciplines have illuminated the process of early bilingualism. First, research on early brain development has shown that infants can learn two languages simultaneously and that the early years are the optimal time to become bilingual (Ramirez & Kuhl, 2017). Evidence from cognitive neuroscience shows that the bilingual brain is more active neurologically than the monolingual brain due to the need to process two languages (Bialystok, 2017). This increased early processing demand is associated with greater control of focused attention and self-regulatory behavior (Conboy, 2013), skills that are associated with enhanced executive function in DLLs. Second, research from psycholinguistics has shown that although DLLs follow a general language trajectory similar to that of monolingual children, their development will demonstrate unique characteristics as a function of learning two languages. These characteristics include language mixing, smaller vocabularies in each language (Bedore, Peña, García, & Cortez, 2005), and differences in the emergence of certain linguistic benchmarks (NASEM, 2017).

A 2017 report by the National Academy of Sciences, Engineering, and Medicine (NASEM), *Fostering the Educational Success of Children and Youth Learning English*, offers a research synthesis on the development and achievement of DLLs from birth to age 21. This consensus study has yielded a comprehensive view on language development, school practices, and educational policies that impact DLLs' growth and school success. It reports four major interrelated conclusions that are central to improving the educational outcomes for DLLs. First, all children are capable of learning more than one language from the earliest months of life and benefit from early exposure to multiple languages. Second, high levels of proficiency in both the home language and English are linked to the best academic and social outcomes. Third, the earlier a child is exposed to a second language, the greater their chances for full bilingualism.

---

### Summary of Findings of NASEM (2017) Report for DLLs 0-5.

*The major findings about DLLs ages birth to five from the NASEM (2017) report include the following:*

- *All young children, if given adequate exposure to two languages, can acquire full competence in both languages;*
- *Early bilingualism confers benefits such as improved academic outcomes in school as well as enhancement of certain cognitive skills such as executive functioning;*
- *Early exposure to a second language—before three years of age—is related to better language skills in second language, English;*
- *The language development of DLLs often differs from that of monolingual children: they may take longer to learn some aspects of language that differ between the two languages and their level of proficiency reflects variations of amount and quality of language input;*
- *The cognitive, cultural, and economic benefits of bilingualism are tied to high levels of competence including listening, speaking, reading, and writing in both languages, e.g., balanced bilingualism at kindergarten entry predicts best long-term outcomes;*
- *DLLs should be supported in maintaining their home language in preschool and early school years while they are learning English in order to achieve full proficiency in both languages;*
- *DLLs language development is enhanced when adults provide frequent, responsive, varied language interactions that include a rich array of diverse words and sentence types. For most DLL families this means they should continue to use their home language in everyday interactions, storytelling, songs, and book readings;*
- *There is wide variation in the language competency among DLLs that is due to multiple social and cultural factors such as parents' immigration status and number of years in U.S., family Socio-Economic Status (SES), status of home language in the community, resources and amount of support and for both languages.*

Source: NASEM (2017). *Promoting the Educational Success of Children and Youth Learning English: Promising Futures*. Washington, DC: The National Academies Press.

Fourth, home language loss is currently the norm for DLLs, particularly once they enter English-speaking ECE settings, which undermine the possibility of full bilingualism and may place the child at risk for unhealthy family relations, including estrangement from their cultural heritage. (See text box for a summary of the NASEM findings for DLLs.)

The NASEM report findings are contributing to an emerging consensus on the elements of effective practices for DLLs. An underlying principle for the effective education of DLLs is early and systematic exposure to English as well as intentional support for home language maintenance and development. Early balanced and intentional exposure to both languages supports early bilingualism, which is important for kindergarten entry and later academic success. Research has identified certain home environment and ECE program features and instructional practices that promote school readiness and help reduce the achievement gap between DLLs and their English-only peers at kindergarten entry.

**ECE teachers must adopt strategies that recognize, value, and integrate the use of DLLs' home languages into classroom practices.**

Home language preservation should be considered a priority for all ECE programs. When very young DLL children are exposed to English, they often start to show a preference for speaking English and a reluctance to continue speaking their home language (Wong-Fillmore, 2001; Oller & Eilers, 2002). ECE professionals and program administrators should know that there are developmental risks associated with the loss of a child's first language. As English constitutes the primary language that DLLs hear outside the home, and it is often the preferred language in community contexts, it is very easy

for DLLs to lose their desire and ability to understand and speak their home language, especially once they are exposed to English in an ECE setting that uses English as the language of instruction. Therefore, ECE teachers must adopt strategies that recognize, value, and integrate the use of DLLs' home languages into classroom practices.

Ensuring exposure to English during the preschool years is also key. Although some preschool DLLs may be fluent in both languages, others will be proficient in the home language but know very little English, have some English conversational language abilities but few academic language skills, or have minimal proficiency in both languages (Páez & Rinaldi, 2006; Place & Hoff, 2011). Recently, several studies have shown that lower levels of English proficiency at kindergarten entry are related to later school difficulties, specifically in English reading (Galindo, 2010; Halle, Hair, Wandner, McNamara, & Chien, 2012). These studies underscore that systematic exposure to English during the preschool years is also important to DLLs' future school performance. Recent research on the amount of time it takes DLLs to become reclassified as fully proficient in English has also found that early proficiency in both the home language and English at kindergarten entry is critical to the process of becoming academically proficient in a second language and may reduce the amount of time it takes to become reclassified (Thompson, 2015; Ansari & Winsler, 2016). Further, Barbara Conboy's (2013) and others' research has led to a consensus that earlier exposure to two or more languages with frequent enriched language interactions leads to the cognitive advantages associated with bilingualism, as the specific languages a child is learning as well as the amount of experience with each language influences how the brain processes each language.

These bilingual benefits have been found across cultural and socioeconomic groups as well as across different language combinations. However, these cognitive advantages depend on the extent to which the child is bilingual (Gordon, 2016). Children who are more balanced in their bilingualism show larger advantages than children who are more dominant in one language. The fact that preschool DLLs enter programs with some proficiency in their home language and are at an ideal age to learn and benefit from learning a second language, that is, English, provides a compelling rationale for designing programs that support both languages.

To summarize, scientific findings confirm that preschoolers have the capacity and, indeed, are neurologically prepared to learn more than one language—and they gain cognitively from managing the linguistic processing required to become bilingual. However, learning a second language should not come at the expense of continued home language development. The research highlights the importance of sufficient exposure to both languages to reap the benefits of bilingualism.

It is important for educators to recognize that there are differences between DLLs and monolinguals. Preschool DLLs seem to show a different pattern of strengths and needs than monolinguals. They are at risk for low levels of oral language development if they don't receive frequent high-quality enriched language learning opportunities in both languages. Their basic mathematical understandings may differ from those of English speakers if their first language uses different language constructs for expressing math concepts such as counting, plurals, grouping, and so forth. They may also excel in certain executive function skills such as cognitive control, and they often demonstrate social-emotional strengths (NASEM, 2017).

In some areas of development, preschool bilinguals show either no differences or slight developmental gaps when compared to monolingual children. For instance, Sandhofer and Uchikoshi (2013) point out that studies have consistently found that bilingual children take longer to recall words from memory. They have slower word retrieval times in picture naming tasks and lower scores on verbal fluency tasks. These findings underscore the need for teachers to understand the challenges a young DLL experiences when processing language, particularly the nondominant language, and the need to allow sufficient time for the child to come up with a response. It is important to give all children sufficient time to respond, but it is critical for young DLLs who are processing language requests in two languages.

In addition, many studies have found that bilingual preschoolers tend to have smaller vocabularies in each language when compared to English-speaking and Spanish-speaking monolinguals. However, a DLL's vocabulary is distributed across two languages; when both languages are considered, their vocabulary size is often comparable to that of monolinguals. As Conboy (2013) has pointed out, "Bilingual lexical learning leads to initially smaller vocabularies in *each separate language* than for monolingual learners of those same languages, but that *total vocabulary sizes* (the sum of what children know in both their languages) in bilingual toddlers are similar to those of monolingual toddlers" (p. 25).

Because vocabulary size is a key goal in preschool and very important to future reading comprehension, this variation in dual language learning is critical for preschool teachers to understand. The difference in DLLs' vocabulary development most often does not indicate language delays or possible learning problems but is a typical feature of early bilingualism. If a preschool child does not know the English word for book, the child may nonetheless understand the concept of a book but know it by a different word such as *libro*.

To sum up, multiple factors are known to affect DLLs' vocabulary growth including similarities between the two languages being learned, the language of schooling, age of acquisition of each language, the child's family socioeconomic status, and the quality and quantity of their exposure to each language. Further, DLLs typically develop vocabulary knowledge in different contexts such as home or school for each of their languages, and the rate of vocabulary development may not be the same for each language (NASEM, 2017; Espinosa, 2015).

Oral language skills, including vocabulary skills, listening comprehension, grammatical knowledge, and expressive vocabulary, have been found to be especially important for DLLs' future reading abilities. Recent research with young Spanish-speaking children from low socioeconomic backgrounds has found that these young DLLs might be at risk for delays in their early literacy development due to their weaker oral language abilities (Espinosa & Zepeda, 2016; Mancilla-Martinez & Lesaux, 2011). This research with dual language learners demonstrates the need to promote oral language development by providing rich and engaging language environments in both languages while at the same time focusing on building early literacy skills. In light of this research, it is essential for preschool programs to recognize the critical importance of oral language and vocabulary development for young DLLs.

Knowledge of linguistically appropriate assessment practices for DLLs is particularly crucial. Valid and comprehensive assessment of young DLLs' development and achievement is essential yet often challenging for ECE professionals (Espinosa & García 2012). Individualized instruction enhances young children's learning opportunities and promotes the important developmental and achievement outcomes necessary for school success. Individualized instruction, however, requires comprehensive, ongoing assessments that are fair, valid, and linguistically, culturally, and developmentally appropriate. Such assessments show educators what DLLs already know and what needs to be taught.

For DLLs, the language in which an assessment is given will determine how well they score as well as the educational services they receive. Because DLLs acquire their knowledge of the world around them through two languages, their language skills will be distributed across both. Therefore, to get an accurate picture of DLLs' language abilities requires assessment in each of their languages. A DLL child may know some words and concepts in one language and others in the second language. Depending on children's experiences and learning opportunities, they most likely will not perform as well as monolingual speakers of either language. This pattern is a typical and usually temporary phase of emergent bilingualism (Paradis, Genesee, & Crago 2011).

DLLs who are assessed only in the weaker language, such as English—as is often the case with early language and kindergarten readiness assessments—will often score significantly lower in language, literacy, math, and basic concepts tasks than their English-only peers (Espinosa & García, 2012). However, their scores may be typical for children who are in the early stages of second language acquisition and may not represent any language delays or be a cause for concern. Therefore, conclusions about DLL children’s developmental progress or need for special services must be based on knowledge about their abilities in both languages as well as on what should be expected of preschool DLLs and how they differ from monolinguals.

Both formal and informal methods are required to ensure appropriate assessments of DLLs (Espinosa, 2015). Initial assessment should include a formal family interview or questionnaire about what languages spoken in the home and by which family members. Other formal child assessments such as the preLAS (Duncan & De Avila, 1985)—a measure of language proficiency—can be administered to individual children to give ECE personnel more specific information about a child’s receptive and expressive language abilities. In addition to formal assessment, ECE teachers can use ongoing informal observational assessment—both structured and unstructured—to monitor a child’s progress and plan appropriate learning activities.

## IMPLICATIONS OF RESEARCH FOR INSTRUCTIONAL PRACTICES FOR DLLS

---

***Unless you believe “in your bones” that having a second language in addition to English is a gift, and not a disadvantage, and diversity is a resource, not a problem to be solved, you are likely to respond to DLL children in ways that discourage the continued use of the home language, especially if you are not fluent in the child’s home language.***

—Espinosa & Magruder, 2015, p.80

The following instructional strategies and recommendations referenced in the NASEM report (2017) are backed by empirical evidence that shows they promote important academic outcomes for DLLs. It should be noted that particular educational approaches will differ based on a program’s language model and its goals and objectives for first and second language development—that is, full dual language models versus primarily English language development with support for home language maintenance.

### ► **Getting to know the children you are teaching**

Before teachers can specifically address instructional goals and strategies for DLLs, they must first get to know the children. They need to gather formal and informal information on their students’ backgrounds and their early language learning experiences as well as abilities, including how much exposure they have to both the home

language and English and how much they use each. During face-to-face interviews with parents, teachers can learn about family values, language preferences, cultural traditions, and the ability to partner actively with teachers in the classroom.

### ► **Instructional supports**

Although common features of high-quality early education described throughout this volume are beneficial for all children, DLLs require additional instructional support. The NASEM report (2017) outlines a number of instructional strategies and enhancements that have been linked to improved achievement for DLLs in early education settings. Because use of the home language while a child acquires English is associated with higher rates of English proficiency (Méndez, Crais, Castro, & Kainz, 2015), ECE staff who use the home language across content areas will help DLLs develop their conceptual knowledge and promote continued development of the home language while they are acquiring English. In addition, if DLLs receive opportunities to develop listening, speaking, writing, and reading skills in both their languages, over time they will demonstrate higher levels of academic achievement in elementary school (Valentino & Reardon, 2015). An ECE program can adopt any of several language models, ranging from full two-way immersion programs to primarily English-language instruction with systematic support for the home language. It is beyond the scope of this chapter to discuss in detail all of the language models possible in ECE settings, but the underlying principle is that DLLs need systematic, intentional exposure to English while also having opportunities to see, hear, speak, and write in their first or home language. If no staff members speak a child's home language, family members or other fluent speakers of the child's language can be recruited to volunteer in the classroom to tell stories, help create print and labeling that can be posted throughout the classroom, identify culturally relevant materials, and possibly even teach all the children a few words of the family's language. Much research has documented the power of honoring and valuing children's home languages in the classroom (NASEM, 2017). DLLs also need instructional adaptations that explicitly bridge what they already know in their home language and what they need to learn in English such as cognate charts, language labeling, and explicit comparisons between the two languages.

One feature of high-quality classrooms that serve DLLs, whether dual language classrooms or primarily English with support for home language, is the monitoring of the amount of time in each language. Supporting DLLs' overall language development requires sufficient time and frequent language interactions in both languages, but ECE teachers often adopt an informal approach that unintentionally results in the dominance of one language over the other. Therefore, continuous monitoring of when, how much, and by whom each language is used is vitally important.

Giving DLLs the definitions of specific vocabulary words in both their home language and English and exposing them to print in a variety of contexts (e.g., storybook reading, daily schedules, and labels on objects) will also assist their

comprehension and oral language skills. Repetition of vocabulary through multiple readings of familiar storybooks and across different activities will help expand their understanding of word meaning. ECE teachers can also help children comprehend and retain new academic vocabulary by targeting three to four words per day, using pictures and visual cues that convey meaning, embedding targeted academic vocabulary in familiar chants and songs, and using physical gestures linked to particular words. These approaches are good practice for all young children, but they are especially helpful for children who do not understand English and cannot be expected to rely solely on oral language input.

Oral language development, which includes a focus on phonological awareness, vocabulary development, listening comprehension, speaking, and narrative skills, is another tool that helps DLLs. Because strong oral language skills are associated with future literacy skills such as narrative production and reading comprehension, young children need ample opportunities in listening and speaking. We now know that most young DLLs learn the code-related skills important to early literacy, such as letter sounds and knowledge of the alphabet, but have a much harder time developing oral language abilities, like extended English vocabulary and grammatical knowledge, that they need to understand complex text (NASEM, 2017). Therefore, daily instruction must provide targeted and responsive opportunities for young DLLs to listen to, comprehend, and review the vocabulary and to practice the skills integral to oral language development.

Language development should not be isolated and restricted to a topic or time of the day but rather embedded in daily interactions and activities. Contingent, responsive interactions that contain increasing levels of grammatical and word complexity with speakers proficient in the second language and adults who help expand a child's language skills during verbal interactions will support English language development. For example, if a child gives a one-word response in the home language to a question posed in English, the teacher should give the child sufficient time to complete the thought in either language, acknowledge the response positively, and provide a response in English that matches the child's level of comprehension. Most experts in early bilingualism recommend that although teachers should stay in one language during a given activity with preschool DLLs rather than switching between languages, they should also ensure that there are enough activities in each language to promote the program's language goals.

Small group activities are also valuable. Like all young children, DLLs need individual attention. However, because DLLs are learning a new language and must process language inputs through two linguistic systems, they benefit from additional time to practice and build both comprehension and production of language. More time spent in small group activities like dialogic reading or vocabulary instruction will allow teachers to individualize interactions with DLLs, informally assess their level of understanding, and probe their language needs. DLLs are often reluctant to participate actively in large group activities, particularly when their English language skills are not well developed. Recent research also demonstrates that DLLs' peers play an important role in their language development (Sawyer et

al., 2018). Most DLLs are highly motivated and eager to interact socially with peers, which gives them opportunities to practice their emerging language skills without adult pressure. Teachers should structure ECE environments and daily schedules with time for both informal (e.g., dramatic play) and formal (e.g., structured partner learning activities) peer interactions throughout the day.

Last, ECE classrooms should reflect the children and families enrolled. Evidence suggests that creating a supportive environment that reflects DLL children’s language and culture will help them feel accepted and welcome, thus promoting positive learning. Displaying pictures and artifacts that represent each family, their home culture, and their family history provides a welcoming and familiar atmosphere. Culturally responsive classrooms have teachers who acknowledge the presence of culturally and linguistically diverse students and create environments in which DLLs feel comfortable, accepted, safe, and intellectually engaged. In such programs, teachers recognize the strengths and needs of their students, convey positive attitudes toward bilingualism, and implement instructional strategies such as those described here that promote early bilingualism and academic achievement. In these ways, teachers create a climate that recognizes the unique characteristics of each child while also setting challenging but achievable goals.

## QUALIFICATIONS OF ECE PROFESSIONALS WHO WORK WITH DLLS

---

If DLLs are to have equitable educational opportunities, an essential element is the qualifications and competencies of the ECE professionals that provide the services. The Institute of Medicine and the National Research Council’s report *Transforming the Workforce for Children Birth Through Age 8: A Unifying Foundation* (Institute of Medicine & National Research Council, 2015) identifies “professionals with regular (daily or near-daily), direct responsibilities for the care and education of young children” as *educators* (p. 27). The quality of these educators has a direct and significant impact on DLLs’ overall development, including their language proficiencies (NASEM, 2017). This section briefly summarizes the recommendations for ECE educators who work with young DLLs.

Currently few states require ECE teachers who work with young DLLs to have specialized training or coursework focused on meeting the needs of such children and their families (Espinosa & Calderon, 2015). The NASEM (2017) report concludes, “The educator workforce, including early care and education providers, educational administrators, and teachers, is inadequately prepared during preservice training to promote desired educational outcomes for dual language learners” (p. 462). For educators working with DLLs, the report recommends a common course of core content that includes the following elements (NASEM, 2017):

- an understanding of language development and the relationship between first and second language development;
- an understanding of the influences of sociocultural factors on language learning;

- knowledge of and ability to implement effective practices for promoting the successful education of DLLs/English learners, including early intervention strategies for DLLs/English learners with disabilities;
- an understanding of assessment instruments and procedures and of how to interpret and apply assessment results for DLLs/English learners;
- development of skills for establishing respectful partnerships with families of DLLs/English learners; and,
- development of skills to advocate on behalf of DLLs/English learners.

In addition, Zepeda (2015), in a paper commissioned for the NASEM report, reviews the research and identifies the following important competencies for people who work with infant, toddler, and preschool DLLs:

- understanding the relationship between early brain development and language development;
- recognizing that switching between languages is a normal part of early bilingualism and not a sign of confusion;
- understanding how to support oral language development in the first and second language;
- recognizing that children’s first language is the medium through which they learn about the values and beliefs of their culture.

**To provide equitable educational services to DLLs, we need an expanded perspective that recognizes their strengths and potential for cognitive, linguistic, and social advantages, not one that views DLLs’ development as “deficient” because of their limited English skills or one that is based on expectations for monolingual English-only children.**

Though there is widespread agreement among bilingual scholars that it takes specialized knowledge and competencies to work effectively with DLLs, very few states address this issue in their ECE teacher preparation programs. Moreover, ECE professional development efforts often fall short, and licensing or credentialing programs rarely include much content focused on second language learning (Espinosa & Zepeda, in press). Generally, at every level of ECE professional preparation and training, expertise on effective pedagogy for DLLs is limited. To provide equitable educational services to DLLs, we need an expanded perspective that recognizes their strengths and potential for cognitive, linguistic, and social advantages, not

one that views DLLs’ development as “deficient” because of their limited English skills or one that is based on expectations for monolingual English-only children. The challenges to including this expanded perspective and DLL-specific knowledge into the complex system of ECE preservice and professional development, although significant, must be addressed through diversification of higher education faculty and ECE workforce development.

## DIRECTIONS FOR FUTURE RESEARCH

---

Substantial research has been done on the capacity of all children to successfully become bilingual, the factors that influence early bilingualism, and the attendant cognitive, linguistic, and social advantages, and there is also an emerging scholarship on effective practices for DLLs. Yet there are still many gaps in our knowledge. The following research topics are derived from the preceding literature review and discussion:

### ***Instruction***

- Which instructional strategies are most effective with different populations of DLLs from a range of linguistic backgrounds, that is, when the languages represented are highly diverse and dissimilar to English, when the proportion of DLLs ranges from few to mostly DLLs, and when DLLs run the gamut with respect to prior English exposure and proficiency?
- How do different language models—e.g., 90-10, 80-10, or 50-50—impact the acquisition of English during the ECE years?
- At what age should young DLLs attending ECE programs be exposed to English, and what is the ideal amount of early exposure?
- What characteristics of teacher-child interactions support improved school readiness?
- How do differential language proficiencies at school entry affect the learning trajectories of DLLs over the course of K-12 education?
- What are the most effective accommodations for early balanced bilingualism and academic success and what and educational enhancements promote it?

### ***Assessment***

- What are the best assessment tools and procedures to accurately capture the strengths and needs of children who speak more than one language? What combination of formal and informal assessments is needed for developmental screening, measuring progress, and accountability?
- How can we develop a profile of normative development for DLLs from a wide range of linguistic and sociocultural backgrounds that guides educational decisions such as whether a child has a developmental disability, is ready for school, or is making sufficient progress?

***Implementation Research***

- What are the most effective ECE teacher preparation and professional development models for teachers serving DLLs?
- What are the core elements and necessary supports for effective implementation of dual language program models, for example, 50-50, 90-10, and 80-20?
- What are the necessary conditions in communities, programs, staff, and schools for successful implementation of a preschool bilingual program?
- What are the barriers to implementing a preschool bilingual language model?

## References

---

- Ansari, A., & Winsler, A. (2016). Kindergarten readiness for low-income and ethnically diverse children attending publicly funded preschool programs in Miami. *Early Childhood Research Quarterly*, 37, 69-80.
- Barnett, W. S., Carolan, M. E., Squires, J. H., Clarke Brown, K., & Horowitz, M. (2015). *The state of preschool 2014: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.
- Bedore L., Peña, E., García, M., & Cortez, C. (2005). Conceptual versus monolingual scoring: When does it make a difference? *Language, Speech, and Hearing Services in Schools*, 36(3), 188-200.
- Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, 143(3), 233-262.
- Billings, E. S. (2009). El alfabetismo y las familias latinas: A critical perspective on the literacy values and practices of Latino families with young children. *Journal of Latinos and Education*, 8(4), 252-269.
- Castro, D.C., Espinosa, L. M., and Páez, M. (2011). Defining and measuring quality 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.
- Center for Early Care and Education Research—Dual Language Learners. (2011). *Early care and education quality measures: A critical review of the research related to dual language learners* (Research Brief 5). Chapel Hill: University of North Carolina, Frank Porter Graham Child Development Institute, Center for Early Care and Education Research—Dual Language Learners.
- Conboy, B. (2013). Neuroscience research: How experience with one or multiple languages affects the developing brain. In L. M. Espinosa (Ed.), *California's best practices for young dual language learners* (pp. 1-50). Sacramento: California Department of Education.
- Duncan, S. E., & De Avila, E. A. (1985). *Pre-Language Assessment Scales*. Monterey, CA: CTB/McGraw-Hill.
- Espinosa, L. M. (2015). *Getting it right for young children from diverse backgrounds: Applying research to improve practice with a focus on dual language learners* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Espinosa, L. M., & Calderon, M. (2015). *State early learning and development standards, policies and related practices: How responsive are they to the needs of young dual language learners?* ( BUILD Initiative report). <http://www.buildinitiative.org/Portals/0/Uploads/Documents/State%20Early%20Learning%20and%20Development%20Standards,%20Policies%20and%20Related%20Practices%20pdf>.
- Espinosa, L. M., & García, E. E. (2012). *Developmental assessment of young dual language learners with a focus on kindergarten entry assessments: Implications for state policies* (Working paper 1). Chapel Hill: University of North Carolina, Frank Porter Graham Child Development Institute.
- Espinosa, L. M., & Magruder, E. S. (2015). Practical and proven strategies for teaching young dual language learners." In Espinosa, L. M. (Ed.), *Getting it right for young children from diverse backgrounds: Applying research to improve practice with a focus on dual language learners*, 2nd edition (pp. 76-113). Upper Saddle River, NJ: Pearson.
- Espinosa, L. M., & Zepeda, M. (2019). Linguistic and cultural diversity: Knowledge utilization in ECE. In B. Wasik & S. Odom (Eds.), *Celebrating 50 years of child development research: Past, present and future perspectives* (pp.75-92). Baltimore, MD: Paul H. Brookes.
- First Five California (August, 2017). *What constitutes high quality early learning experiences for California's young dual language Learners* (Unpublished working paper).

## CHAPTER 6 ADDRESSING EQUITY IN THE ECE CLASSROOM: EQUAL ACCESS AND HIGH QUALITY FOR DUAL LANGUAGE LEARNERS

- Fuller, B., Bein, E., Kim, Y., & Rabe-Hesketh, S. (2015). Differing cognitive trajectories of Mexican American toddlers: The role of class, nativity, and maternal practices. *Hispanic Journal of Behavioral Sciences, 37*(2), 139-169.
- Galindo, C. (2010). English language learners' math and reading achievement trajectories in the elementary grades. In E. E. García & E. C. Frede (Eds.), *Young English language learners: Current research and emerging directions for practice and policy* (pp. 42-58). New York: Teachers College Press.
- Gordon, K. R. (2016). High proficiency across two languages is related to better mental state reasoning for bilingual children. *Journal of Child Language, 43*(2), 407-424.
- Halgunseth, L., Jia, G., & Barbarin, O. (2013). Family engagement in early childhood programs: Serving families of young dual language learners. In L. M. Espinosa (Ed.), *California's best practices for young dual language learners* (pp.119-171). Sacramento: California Department of Education.
- Halle, T., Hair, E., Wandner, L., McNamara, M. & Chien, N. (2012). Predictors and outcomes of early versus later English language proficiency among English language learners. *Early Childhood Research Quarterly, 27*(1), 1-20.
- Hammer, C.S., Komaroff, E., Rodriguez, B., Lopez, L., Scarpino, S., & Goldstein, B. (2012). Predicting Spanish-English bilingual children's language abilities. *Journal of Speech-Language-Hearing Research, 55*(5), 1251-1264.
- Institute of Medicine & National Research Council. (2015). *Transforming the workforce for children birth through age 8: A unifying foundation*. Washington, DC: The National Academies Press.
- Mancilla-Martinez, J., & Lesaux, N. (2011). The gap between Spanish-Speakers' word reading and word knowledge: A longitudinal study. *Child Development, 82*(5), 1544-60.
- Méndez L. I., Crais E., Castro D., & Kainz K. A. (2015). Culturally and linguistically responsive vocabulary approach for young Latino dual language learners. *Journal of Speech, Language, and Hearing Research, 58*(1), 93-106.
- National Academies of Sciences, Engineering, and Medicine (NASEM). 2017. *Promoting the educational success of children and youth learning English: Promising futures*. Washington, DC: The National Academies Press.
- National Association for the Education of Young Children. (2019). *Advancing equity in early childhood education: A position statement*. Washington, DC: NAEYC.
- National Institute for Early Education Research. (2017). *The state of preschool 2016: State preschool yearbook*. New Brunswick, NJ: NIEER.
- National Institute for Early Education Research. (2018). *The state of preschool 2017: State preschool yearbook*. New Brunswick, NJ: NIEER.
- Office of Head Start. (2015). *Head Start early learning outcomes framework: Birth to five*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.
- Office of Head Start. (2017). *Home Language Support*. <https://eclkc.ohs.acf.hhs.gov/culture-language/article/home-language-support>.
- Oller, D. K, & Eilers, R. E. (2002). *Language and literacy in bilingual children*. Clevedon, UK: Multilingual Matters.
- Olsen, L. (2010). *Reparable harm: Fulfilling the unkept promise of educational opportunity for California's long-term English learners*. Long Beach, CA: Californians Together.
- Páez, M., & Rinaldi, C. (2006). Predicting English word reading skills for Spanish speaking students in first grade. *Topics in Language Disorders, 26*(4), 338-350.

- Paradis, J., Genesee, F., & Crago, M. (2011). *Dual language development and disorders: A handbook on bilingualism and second language learning*. Baltimore, MD: Brookes Publishing.
- Park, M., McHugh, M., Zong, J., & Batalova, J. (2015). *Immigrant and refugee workers in the early childhood field: Taking a closer look*. Washington, DC: Migration Policy Institute.
- Park, M., Zong, J. & Batalova, J. (2018). *Growing superdiversity among young U.S. dual language learners and its implications*. Washington DC: Migration Policy Institute.
- Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Development, 82*(6), 1834-1849.
- Ramirez, N. F., & Kuhl, P. (2017). The brain science of bilingualism. *Young Children, 72*(2), 38-44.
- Sandhofer, C., & Uchikoshi, Y. (2013). Cognitive consequences of dual language learning: Cognitive function, language and literacy, science and mathematics and socioemotional development. In L. M. Espinosa (Ed.), *California's best practices for young dual language learners* (pp. 51-89). Sacramento: California Department of Education.
- Sawyer, B., Atkins-Burnett, S., Sandilos, L., Hammer, C. S., Lopez, L., & Blair, C. (2018). Variations in classroom environments of preschool children who are low income and linguistically diverse. *Early Education and Development, 29*(3), 398-416.
- Small, M. F. (2002). *Kids: How biology and culture shape the way we raise young children*. New York, NY: Anchor Books
- Thompson, K. D. (2015). English learners' time to reclassification: An analysis. *Educational Policy, 31*(3), 1-34.
- Valentino, R. A., & Reardon, S. F. (2015). Effectiveness of four instructional programs designed to service English learners: Variation by ethnicity and initial English proficiency. *Educational Evaluation and Policy Analysis, 37*(4), 612-637.
- Wong-Fillmore, L. (2001). When learning a second language means losing the first. *Early Childhood Research Quarterly, 6*(3), 323-346.
- Zepeda, M. (2015). *The early childhood workforce for dual language learners: What do we know and where do we need to go?* Unpublished manuscript, California State University, Los Angeles, Department of Child and Family Studies.



**SECTION 2, CHAPTER 7**

**VIGNETTE: BUILDING A  
HIGH-QUALITY PROGRAM—  
THE BOSTON PUBLIC SCHOOLS  
EXPERIENCE**

---

Jason Sachs, Ed.D., Boston Public Schools



***I usually start my presentations with an image of a carousel with one horse taking off to acknowledge that our work is heavily situated within our own contexts, stakeholders, and resources. My story starts in a public school district, but I suspect that readers will come from many places—for example, state education departments, city governments, local agencies, and school districts. The Boston Public Schools’ early childhood education program, which I lead, often spans multiple domains—academic, operational, budgetary, prekindergarten and kindergarten, early elementary, and the like. I hope that readers will become like the horse breaking free—taking what is useful for their own contexts—and that this article will help your work as you set out to build or improve your own preschool systems and partner with your own public schools.***

—Jason Sachs

The story of the early childhood initiatives undertaken by Boston Public Schools (BPS) starts long before I arrived. Boston was home to the first public school in America and also the first kindergarten. By the time I joined BPS over a decade ago, six early education centers were running full-day programs for prekindergarten up to first grade and were headed by principals who were outspoken leaders in early childhood education in both the district and in the city. The district had run half-day programs for 4-year-olds in the 1990s, but it cut that program to create resources for full-day kindergarten for all 5-year-olds. In 2005, Mayor Thomas Menino and Superintendent Thomas Payzant, both veterans in their jobs, decided it was time to serve 4-year-olds again, and almost overnight they created a universal prekindergarten program. The program was to be delivered in schools in the BPS system; it would be free for all, and teachers would be paid on the same scale and receive the same benefits as K-12 teachers and be subject to the same educational and certification requirements (e.g., they would need to earn a master’s degree within 5 years). After this momentous decision, I was hired to lead the newly created Department of Early Childhood. The mayor and the superintendent at the time had each been in his position for almost a decade and had provided steady leadership and support, which turned out to be very important to the success of the program.

Before I took the job with the BPS, I worked for the Massachusetts Department of Education’s Early Learning Services, which oversaw the distribution of \$128 million in funds for programs from birth (family support and home visiting) through kindergarten. The work I did at the state level influenced how I saw policy tools such as accreditation from the National Association for the Education of Young Children (NAEYC), quality enhancements, professional development (PD), home visiting, evaluation, budgeting, and collaboration. It also influenced my views on management. For example, I believe that strong leaders act as facilitators, pose problems, listen, and usually speak last. I also learned how to navigate in a large bureaucracy where leaders, politics, and priorities are constantly changing. A statewide view showed me that the leadership of public schools, Head Start, and community-based programs varies from community to community, as does the quality of the services these programs offer.

Other lessons center around the importance of local collaboration, accountability, relevant real-time data, the nature of funding mechanisms (grants versus child subsidies), and capacity building. I also learned that things can be both created and dismantled very quickly, so it is important to build systems and structures that can withstand changing priorities.

Taking what I learned from the state and before, I spent 5 years working in Boston for a large child-care agency run by Douglas Baird, an outspoken leader for early education reform. Working in and for a community-based organization gave me the opportunity to see the fiscal challenges created by low state reimbursement rates for low-income child-care subsidies funded by the state and federal governments, a subject that had been an interest of mine since my PhD days. My dissertation focused on the consequences low-quality early education programming on students' outcomes. Once I knew the harm that seemingly well-intentioned policies were causing, my life's trajectory was set.

## BUILDING SYSTEMS: THE WORK OF THE BPS DEPARTMENT OF EARLY CHILDHOOD

---

To build systems, you have to think in terms of a 3- to 5-year arc, knowing that you are going to have to make tactical shifts along the way. The choices you make should be strategic: the goal should be services that are both needed and possible to secure. It took us 6 years, for example, to implement a kindergarten curriculum across the district and almost 9 years to meaningfully link our curriculum to families. It was only in our 12th year that we were able to introduce a formative assessment system based on observation and documentation. In this chapter, I share with you the larger projects we did along the way, many of which persist to this day in modified forms. For example, we decided to use a centralized pre-K curriculum but have since rewritten it, and we have also developed a kindergarten to second-grade curriculum that draws on some of the same instructional practices that we use in the pre-K program.

**To build systems, you have to think in terms of a 3- to 5-year arc, knowing that you are going to have to make tactical shifts along the way. The choices you make should be strategic: the goal should be services that are both needed and possible to secure.**

## BASIC FACTS ABOUT THE BPS EARLY CHILDHOOD PROGRAM

---

Under the program developed by the mayor and superintendent in 2005, K1 (our pre-K program for 4-year-olds) is the same as any other grade in the district. The only difference is that there is a full-time paraprofessional in every classroom. Our staff to student ratio is 1:11. The program operates on a normal BPS school-day and school-year schedule, and enrollment is based on a lottery system. We currently serve roughly 55% of all 4-year-olds in the city and have a waitlist of well over 1,000. The BPS pays for the services out of its own budget. The per-pupil cost is about the same as for kindergarten or fifth-grade students. Though the cost of the program to the district is reported to be around \$10,000 per pupil per year, the true cost is more like \$17,000 per pupil per year, owing to building maintenance and salaries for principals and support teams.

## CREATING A DEPARTMENT OF EARLY CHILDHOOD

---

You can't really go anywhere with a group of people if you don't know where you are going and cannot convince the people with you that they want to go as well. That's why we developed a mission statement for the BPS Department of Early Childhood in 2006. The department aims "to ensure that principals, teachers, paraprofessionals and school support staff have the knowledge, skills and resources they need to provide a high-quality early education experience for all students," and its "expectation is that all children will become internally driven and self-motivated learners and will be able to read, write and communicate effectively by third grade."<sup>1</sup> Lately, I have been thinking that we should change "communicate effectively by third grade" to "communicate effectively and with passion by third grade." We are also contemplating adding "and compute" after "communicate" to acknowledge the importance of math skills.

As a team, we have grown from two to 24 people, and we now oversee the citywide universal pre-K program and have curriculum oversight for preschool through second grade. Eighty percent of the staff are program developers, that is, coaches, and they are a large part of our success. They are the main body of our staff and spend at least 50% of their work time in classrooms. Coaches are in a different union from the BPS teachers, so they can also provide evaluation assistance to principals. However, because the relationship between a teacher and his or her coach is nonevaluative, we use a different coach to evaluate the teacher. Coaches in general have master's degrees and are paid as much as BPS teachers or more. We have four managers—one for NAEYC accreditation, one for the universal pre-K program, one for budgets and work plans, and one for research and grant writing. Having the majority of our staff in classrooms makes us aware of the real impact of our work. Schools and classrooms are dynamic places, and we have to compete with other school and district priorities, so having coaches lead most of our work shows us what is both needed and realistic.

---

<sup>1</sup> <https://www.bostonpublicschools.org/earlychildhood>

We have a blended funding model that secures us resources from state, federal, and private entities. Forty percent of our staff are paid with outside grants, and the district covers the rest. Having outside funding sources is helpful for two reasons: it allows us to innovate and be flexible (city/state funds usually have to be used in specific ways), and it also holds us accountable to our private funding sources, which often require evaluation data. I have the unique opportunity to combine BPS general funds with private funding dollars. While the resources have priorities and associated accountability, there is enough tolerance in the funding that I am allowed to start new projects and also shift resources when needed. For example, both times we launched a curriculum pilot, more schools applied than we anticipated; rather than limit them, we were able to accommodate them. This decision, though it drained more resources, allowed us to serve more students in real time than if I had been constrained by the original design of the funding partner.

We are a productive group. We like to complete tasks and move on to the next large project, because many other areas—special education, learning assessments, dual language considerations, toxic family stress—need our attention. We use work plans and the evaluation system to help us focus on our priorities. We usually spend the end of May through July celebrating, analyzing our challenges, and then planning and prioritizing our work for the next school year. From August to October, we create and enact implementation plans, and from November through April we focus our efforts on schools and have monthly staff meetings that alternate between PD and coaching calibration. Grade and project teams meet weekly. This process allows us time both to reflect by providing natural break points during which to assess our progress and to productively struggle in the field where day-to-day progress seems slower.

Staff are also allowed to spend up to 20% of their time on a goal that they feel will effect change, for example, linking curriculum to families, incorporating “beautiful stuff” into the curriculum, or connecting with outside partnerships. Many of the innovations—and, subsequently, strategies—of the department come from staff members embracing their passions in this way.

## COACHING AND PROFESSIONAL DEVELOPMENT

---

We have tried a variety of coaching models, with ratios as low as one coach to eight teachers and as high as one coach per 20 (more of a grade-level team focus). What we have learned is that coaching is most effective when the teacher wants to change and that the strategies we use need to be differentiated based on a teacher’s knowledge level and how committed the school or program is to change. Loosely, teachers fall into three categories: those who need to be evaluated out; those who can grow with coaching through biweekly visits; and those who do not need much coaching or who attend seminars with peers. We have also had to work carefully on what kinds of coaching goals we pursue, focusing, for example, on curriculum knowledge transference rather than good early childhood practice because the former is much clearer and easier to coach and measure through fidelity scores.

Our PD model is relatively standardized and linked to coaching. That is, for the most part, if you attend the PD you get coaching, as the two are linked in scope and sequence. In the summer we take 3 to 5 days to introduce our curriculum to new teachers, and then we have monthly seminars—run like graduate school classes with smaller cohorts—to support their curriculum instruction. Videotaping, teacher documentation of student work, and webinars are becoming more common in our practice, and we have much more room to grow in these areas.

The lion’s share of our PD focuses on first setting the table—getting teachers to understand their curriculum and the “whys” underneath it, and then getting them to reflect about who they are teaching and how differentiate their instruction. Though we focus on curriculum fidelity, we view it as “a tool, not a rule.” We know that strong teachers will need to make adjustments along the way to meet the diverse needs of their classrooms. The rub is getting them to make choices based on what facilitates learning versus what is easier to manage.

## WHO AND HOW WE HIRE

---

At BPS we work hard to hire coaches who represent the early childhood field. Hence we hire teachers from community-based programs, district literacy coaches, directors of education programs, and principals. Below are sample questions we use for hiring staff. These questions address the depth of knowledge our coaches need and underscore our commitment to the population we are serving and the importance of early literacy.

- What is your approach to collaboration? What do you expect of others? What do you do when your perspective differs from the perspectives of others?
- Please describe any experience you have working with low income, culturally diverse children and families. Include your experience working with children whose first language is not English or children with special needs. What do you draw from these experiences that would help you as a program developer or coach?
- What does developmentally appropriate practice mean to you? Why is it important and how do you incorporate this pedagogy into your practice?
- Talk about your experience teaching early literacy. What approaches have you followed and what resources have you relied on? What do you believe are the critical components to building and supporting strong early readers and writers?
- What is your approach to integrating content areas? For example, how do you see connections between literacy and science or math and social studies?
- Describe your experience with coaching or mentoring teachers (for example, observing, planning, modeling, and debriefing lessons). What is your approach to moving a teacher’s practice?

- How do you advise a teacher who recognizes the interest of an individual child or group of children that strays from the path of the established curriculum? How might you respond to this tension?
- How would you develop a relationship with the principal/administrative staff to facilitate your success as a program developer? Please give examples of specific things you would do.
- Please talk about your experience and comfort in providing PD for teachers and administrators. What ideas do you have about the most effective ways to pass on professional knowledge?

## SELECTION CRITERIA FOR SELECTING CLASSROOMS IN BPS

---

We had to establish some basic selection criteria based on supply and demand, quality of facilities, and school capacity to determine where to place classrooms:

- We did not want to create a single early childhood strand, as teachers work better in pairs.
- We had to place as many pre-K classrooms in schools as there were kindergarten classrooms.
- We had to place classrooms on first or second floors with bathrooms within 40 feet of them to meet NAEYC standards criteria.
- We had to put classrooms in schools where there was demand.
- We had to look at the choice of where to put classrooms through an equity lens of who would get access.
- The school needed to have stable leadership in place to take on more students.

In the early days, we grew from serving roughly 400 students in 30 mixed inclusion classrooms in 2005 to serving over 2,500 4-year-olds in over 150 classrooms in more than 70 elementary schools by 2010.

## RESEARCH AND EVALUATION: THE ROLE OF DATA IN THE PROCESS OF CHANGE

---

In this section, I offer a brief history of the Department of Early Childhood’s use of data and evaluations to guide program and practice. The use of research and data to drive change by the department got off to what many would consider an inauspicious start.<sup>2</sup> After just 2 years of operation, it hired an outside research firm to measure the quality of its classrooms. The findings were displayed prominently on the first page of the *Boston Globe*: “Boston Preschools Falling Far Short of Goals,” the headline read, with the story noting that “the city’s public preschool and

---

<sup>2</sup> This section was written in collaboration with Christina Weiland, and parts of it appear in a book by Betty Bardige, Megina Baker, and Ben Mardell (2018) about the Boston Public Schools and its early childhood efforts. Chris has collaborated with our department on almost all data and evaluation work. She started out as an intern and is now an assistant professor at the University of Michigan. Having a researcher along every step of the way has strengthened the program immeasurably (pun intended).

kindergarten programs are hobbled by mediocre instruction” (Jan, 2007). The findings could have jeopardized the whole endeavor of public preschool in Boston, by creating both a “see, we told the BPS they couldn’t do this” mindset and mistrust among teachers. On both counts, we survived. We did so thanks to strong leadership from the mayor and superintendent and by communicating directly with teachers and listening to the “why” behind the findings. For example, teachers said that they did not have strong curriculums, that their principals did not let them teach in developmentally appropriate ways, and that they spent too much time assessing students. The 2006 findings, however, played a large role in shaping our strategic plan and taught us that the BPS, the school committee, and the city council can tolerate negative findings, which allowed us to continue to evaluate and revise our work going forward.

## RESEARCH AND EVALUATION IN THE DEPARTMENT OF EARLY CHILDHOOD, 2006–2017

---

Over the course of the department’s history, we have collected and used data in a variety of ways. Table 1 illustrates the data types we use, how frequently these data are collected, their purpose, and how we use them to drive change. The table is purposely broad so as to give a gestalt understanding and not overwhelm the reader with information pertaining to every data type and every wave of data collection.

The outside team produces a report with central findings and also a dataset for the district’s use. We use their findings to help the department make programmatic and district policy decisions and also to perform our own analyses, often linking their dataset to other sources of data available internally, such as administrative data on program demographics. Partnerships with outside researchers bring an additional perspective on what the results mean and provide more objectivity. Importantly, we are careful in our contracts with outside firms to retain full access to the identified data so that we are not limited in the kinds of internal research that are subsequently possible.

### ► Multipurpose data use

As Table 1 illustrates, the Department of Early Childhood uses data for a variety of purposes, such as identifying systematic weaknesses across classrooms and targeting PD accordingly. For example, classroom quality data collected in 2010 in prekindergarten and kindergarten revealed that although the program had the highest instructional quality of any large-scale prekindergarten to date (Weiland, Ulvestad, Sachs, & Yoshikawa, 2013), teachers were not doing enough to support children’s conceptual development. Professional development was then modified to target best practices in this area. We also created a teacher-friendly template that displayed each teacher’s results compared to district averages and areas for growth. Coaches worked with teachers to help them understand the implications of their scores for their practices.

Table 1. Summary of types of data collected, frequency of collection, and use

Data source	When collected	Purpose	Use
Classroom quality and curriculum fidelity observational scores	About every 2 years	Changes as program evolves; in 2012, for example, data collection focused on K-2 due to concerns about quality of education after prekindergarten	To determine program gaps, needs, and strengths; to guide professional development (PD) and programmatic decisions
Administrative data	Ongoing	To track important programmatic data like child attendance, enrollment, demographics as well as teacher education, certification, and experience	To answer questions about programmatic use and take-up; to describe the BPS population and how it changes over time. These data also are used as control variables in analyses, reducing participant burden
Teacher surveys	About every 2 years	To gather richer data on teacher background, experience of PD, and opinions/desires related to current offerings	To understand teacher population in more depth; to guide PD and programmatic decisions
P-2 child early reading skills and prekindergarten vocabulary	3 times per year (assessed by teachers)	To monitor children’s early literacy and language skill development and to identify supports as needed	To describe the BPS population; to draw on as outcomes in evaluation studies
Broader set of child outcomes	When external funding is available or when a research study under way requires them	To examine children’s levels and growth on a broad set of important outcomes (math, executive function, socioemotional skills)	To describe the BPS population; to draw on as outcomes in evaluation studies

Data are used to link children’s learning to their program experiences. For example, BPS elementary schools vary in how mixed they are in their income demographics. At some schools, nearly all children come from low-income households, while others have approximately equal representation of students from higher- and lower-income backgrounds. Our department was interested in what effect this demographic variation would have on the pre-K program. We believed that because of the way preschool classrooms are structured, children spend a lot of time interacting with each other, and therefore that children learn a lot from each other; we also believed that higher-income children, on average, come to school with stronger language skills and more world knowledge than their lower-income peers. At the time, Harvard Graduate School of Education researchers Christina Weiland and Hirokazu Yoshikawa took up this question and examined whether the proportion of low-income peers was related to children’s gains in their prekindergarten year. They found that having more mixed-income peers (versus only low-income peers) did predict gains in children’s vocabulary skills during prekindergarten (Weiland & Yoshikawa, 2014). These results did not drive a policy change; BPS children are assigned to schools via a centralized choice system. But it did enhance the department’s understanding of what drives children’s gains in early childhood classrooms, and it contributed to conversations in the design of Boston’s mixed-delivery universal pre-K system.

The mixed-income peers study was published in a peer-reviewed academic journal; feedback from peer reviewers helps us make our work more rigorous and more credible. However, more often than not, the work we have done with data sources in Table 1 has not been usable for studies in peer-reviewed journals. The available data are not always complete enough or able to capture the story fully enough to meet these journals’ high standards.

However, the department has been able to make good use of its data internally. For example, in 2010, the district faced a decision regarding whether to continue to offer a summer reading program to kindergarten and first-grade students and whether to extend the program to incoming prekindergarten students. The district was well aware of research showing that low-income children commonly experience summer learning loss (Entwisle & Alexander, 1992) and that high-quality summer enrichment programs are effective in combating this problem (Borman & Dowling, 2006; Jacob & Lefgren, 2004). In late fall 2010, within the structure of our research partnership, we identified key data from the summer 2009 district summer program that could guide the decision (which children chose to attend the program, attendance data, and student outcome data) and the key research questions.

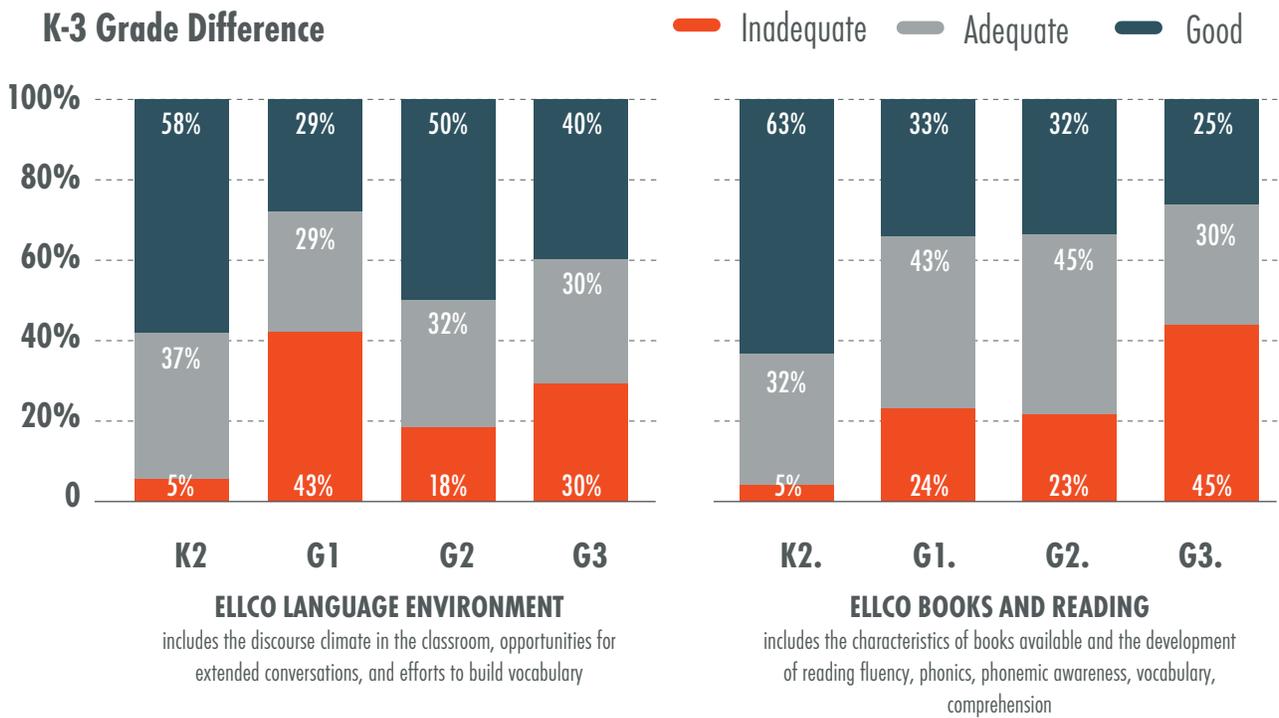
The challenge in answering the research questions rigorously was that students had selected into the program, and so any results, positive or negative, could have had to do with the students themselves and not the program. The research team decided to create two quasi-experimental control groups to increase study rigor: one group was made up of students who applied to the program but did not attend and the other was made up of students attending the same schools as summer-program attenders. Analyses showed that program attendance was strong—80% of students had attendance rates of 73% or higher. The program also reached children more in need of help than their peers; participants had lower literacy skills than their peers prior to the program and were

significantly more likely to have previously repeated a grade. Students who attended the program had stronger post-program literacy skills scores than did children in either of the two control groups. On the basis of this evidence, along with feedback from teachers involved in the program, the district decided not only to continue to offer the program but to offer it to incoming prekindergarten students as well. The program has evolved over time but continues to be offered to young Boston students each summer.

► **Data on fadeout**

“Fadeout” is a hot topic for the field and merits some discussion. Our data are mixed. We definitely see a decline in student gains from pre-K to third grade, but the impact of the BPS’s pre-K program is still significant and substantial. In addition, we still see a gap between black and white students. Our reading fluency (as measured by the DIBELS) data also demonstrate that children who attend K1 score better than students in other pre-K settings and that fewer of them slip into the at-risk category between kindergarten and second grade, so K1 attendance definitely provides some insulating. That said, our data on instructional quality reveal that first through third grade instruction needs improvement, much like preschool and kindergarten did in 2006 (see Figure 1), and hence we have shifted our focus there.

Figure 1. Differences in quality of literacy instruction K-3 (2012).



Source: Department of Early Childhood, Boston Public Schools.

## KEY LESSONS

---

From over a decade of work connecting research to practice, we have drawn a set of key lessons that may be of use to other programs.

*First, there are natural tensions in a research-practice partnership.* Rigor and timeliness often conflict; careful studies can take years, while policy and practice decisions are often made in a matter of weeks or months. As one example, around 2010, a critical decision the district faced was whether to pursue NAEYC accreditation for all district elementary schools. This accreditation process is intended to improve program quality by ensuring that participating early childhood programs meet a set of 10 program standards focused on four main domains: children, teachers and staff, management and administration, and family and community relations. Though NAEYC accreditation is widely considered a marker of quality by the early childhood field, studies have produced limited empirical evidence that it has positive effects on classroom quality and child outcomes (Minnesota Department of Human Services, 2005; Whitebook, Sakai, & Howes, 1997). Accordingly, in 2008, using available district data, we examined whether undertaking accreditation was associated with higher classroom quality in the group of early adopters of the approach in the district. Importantly, schools had selected into accreditation, and the level of rigor we would have preferred was not possible in time to contribute to the district’s decision-making process, but we found that NAEYC accreditation was associated with meaningful improvements in classroom quality (Sachs & Weiland, 2010). The district subsequently used the results of this analysis as one piece of evidence in making its decision to expand NAEYC accreditation to more district schools. Analyses in 2010 and 2015 also examined the role of NAEYC accreditation in the district; the 2015 results led to a shift in NAEYC work that emphasized cognitively demanding tasks for students.

Some questions are too academic in the department’s view; that is, they might benefit the field but not the department. It turns down ideas from Weiland and others that fall into this category if they represent a burden without benefit for the district. Conversely, sometimes the department has had a question or a “need to know” that is either not of interest to academics or not publishable. Weiland and her team have generally taken these on just the same; their view is that to be good citizens and partners and to learn as much about the district as possible, it is important to address them. Finally, a common issue in our work has been that available funders are willing to heavily fund either the research or the program but not both. Research-practice partnership usually requires both, and managing this issue has meant cobbling together sources of support as best we can.

*Second, planning matters.* In September 2007, after 3 months of working with the department, Weiland prepared a memo that included a list of all data collected by the district relevant to the department, study designs that could be appropriate for answering the department’s questions, and an overview of what external funding would be required to collect other types of data. This early exercise—shared and discussed with the department and the BPS director of

research—helped create a strategic plan for the kinds of questions our research partnership would address and when. A key question, for example, was whether the program was ready for an impact study and what funding would be available to carry it out. In accordance with the literature, we jointly determined that 2 years after the implementation of the district’s curricula and biweekly coaching was a good time to determine whether the new model was working. The subsequent study—funded by the Institute of Education Sciences—showed that the model had the largest impacts of any large-scale prekindergarten program to date. These impacts were apparent in both outcomes directly targeted by the program—language, literacy, math, and socioemotional skills—and in a domain that was not directly targeted (executive function) but that is developmentally linked to growth in other domains (Weiland & Yoshikawa, 2013). It was critical that this evaluation was conducted when the program was ready and not before the new changes had had time to take root. A research strategic plan also helped us to be clear about which data would be used for continuous quality improvement and how, as well as how the research and data fit together.

*Third, what you don’t do is as important as what you do.* Importantly, we collect less data than many programs do, particularly teacher-collected data. The department’s philosophy is that teachers should focus on teaching, and it has pushed back against state requirements for teachers to collect data via the formative assessment systems used in most pre-K programs nationally. Weiland reviewed the literature on these systems for the department, and she concluded that there is very little rigorous evidence they provide reliable, valid data or that they change teachers’ practice. Such systems require teachers to collect lengthy data on every child in their classrooms, several times a year, and they generally require paying an administrative per child fee to the licensing company. Instead, we have relied on a sampling approach and limited teacher-collected data as well as short direct assessments of child language and literacy that use well-validated, reliable measures.

*Fourth, data helps you work smarter.* I opened this section by recounting the inauspicious beginning of data use in the Department of Early Childhood that the scary headline on the front page of the *Boston Globe* broadcast to the community. Those very public results caused the department to slow down the pace of its expansion and invest in quality. The next time that it attempted something so ambitious as launching a preschool program, it had learned to build in data and careful piloting from the beginning. Specifically, in 2012, the department was asked to expand its model to community-based preschools in Boston. Accordingly, it carefully built in a pilot of its model in this new context and also conducted a pilot study that included observational quality measures, surveys, and interviews of key stakeholders. After 2.5 years, the results were disappointing. While quality initially increased after coaching and curricula were implemented in the first 1.5 years, these gains were not sustained, and the quality of the community-based organizations remained lower than that of BPS classrooms (Yudron, Weiland, & Sachs, 2016). The pilot study identified six barriers that contributed to implementation failure, including lack of common planning time, teachers’ retention of old curricula, teacher attrition from community-based organizations, too many 3-year-olds in a program targeted to 4-year-olds, and no start time for instruction.

These barriers are being addressed—that is to say, data are helping us get smarter. The department capped the number of 3-year-olds allowed in each classroom to approximately five out of 20 students, standardized the pay increases across community-based organizations so that participating lead teachers in them receive salaries equivalent to those of BPS prekindergarten teachers, and required common planning time. The department also modified the PD it offers to community-based organizations to better incorporate their teachers into district training. Another research team (Abt Associates) is evaluating this new model and expansion effort and sharing data with the department. Findings from the first year of implementation were encouraging, and research continues (Checkoway, Goodson, Grindal, & Hofer, 2017). The pilot project and its associated research components have operated as intended in this respect—that is, as part of a continuous quality improvement system—despite somewhat disappointing overall quality changes in the organizations in the pilot project. In our view, improving preschool nationally requires more such careful program piloting and research to pinpoint specific, practical barriers to program quality improvement.

*Fifth, it is important to create strategic plans, and to stick with them.* Strategic plans are very effective, as they let people know what you are trying to do and how they can help. I have had many, many bosses and partners come and go in 12 years. Having a clear strategic plan with a roadmap and deliverables of what you have done and what you want to do is critical. As part of this process, you should collect data and make adjustments along the way. The data will challenge you, but the data will also provide opportunity. As part of our approach of using data to inform the program, we have created two strategic plans; the first lasted 10 years, and the second is set for 5 years. For us, creating a strategic plan with an embedded holistic theory of change is critical. Prioritizing how we should spend our time and identifying what we think are the effective strategies both help to build consensus and to provide direction for the staff. They also help to orient new staff, leadership, funders, and other stakeholders and allow them to get to know what we are doing.

I spend much of my time setting up structures and finding resources to get the work done. On my end, I usually set up a new project—such as Boston K1DS (which was subsequently supported by a federal preschool expansion grant and is now a city-funded universal pre-K program), a first- and second-grade curriculum, an Institute of Education Sciences longitudinal study, or, most recently, a childhood observational assessment—and then once it's up and running I will move on to the next. Our most recent theory of change is that all children will become internally driven learners, able to read, write, reason, solve problems, and communicate effectively by third grade, and that the BPS will close the achievement gap if we can:

- align our work with the BPS vision, implementation plan, and instructional vision;
- expand the early childhood vision to early elementary grades (first to third);
- use data to consistently improve our curriculum, PD, coaching, and assessments;
- target PD and coaching as a way to make specific changes in instructional practice;
- collaborate with teachers, instructional leaders and other departments;
- build capacity for high-quality pre-K in community-based organizations;
- expand out-of-school time programming to support working families; and
- leverage partnerships to sustain our capacity and share our findings.

Our first strategic plan focused on establishing early childhood systems in the BPS, while the second one is focused on a system to support greater expansion into community-based programs for preschool and for altering the first- and second-grade curriculum. Since our current administration is more aligned with approaches centered on coherence building and instruction and collaboration, we are spending more of our time thinking about how to capitalize on departmental interdependence so that we aren't doing the work all on our own.

*Sixth, the curriculum needs to keep pace with the students.* One of my big takeaways from this job is that even if you run a high-quality pre-K program with strong results, you will lose momentum in student gains if it doesn't keep up. Our curriculum history is robust:

- In 2006, we selected Open the World of Learning (OWL) and Building Blocks.
- In 2010, we wrote the Focus on K2 curriculum.
- In 2012, we re-wrote the Focus on K1 curriculum.
- In 2014, we worked with Nonie Lesaux and the Harvard team and to write Focus on First Grade.
- In 2018, we completed our rewrites of Focus on First and Second Grade.

The math curriculum continues to use Building Blocks, and TERC<sup>3</sup> Investigations and is taught discretely.

Our curricula have several core instructional practices that are threaded across the grades. They all have daily expectations and follow a scope and sequence. Common P-2 instructional practices include:

---

<sup>3</sup> Formerly known as Technical Education Research Centers

- facilitating discourse and feedback<sup>4</sup>
- experiential learning across disciplines<sup>5</sup>
- consideration of variance in development, processes, and perspectives<sup>6</sup>
- promotion of active agency and autonomy<sup>7</sup>
- documentation of teaching and learning

We purposely aligned this work with the district’s essential practices to allow administrators to see the connections between early childhood practices and district initiatives. In addition, we have aligned the practices with the Classroom Observation Tool (CLASS) and with district’s teacher evaluation system. The curricular components we use to facilitate these instructional practices include:

- centers (called “studios” in later grades)
- thinking and feedback, a protocol for sharing work in centers
- theme (4 to 6 units per grade)
- interdisciplinary topics in science and social studies that are literacy focused
- core read-alouds that are read multiple times
- vocabulary development
- culminating projects
- phonics programs (kindergarten to second grade)
- storytelling and story acting
- literacy centers that are dedicated to small group literacy work
- discrete math time using Building Blocks and TERC Investigations

---

<sup>4</sup> <https://depts.washington.edu/cqel/PDFs/DickinsonTeacherChildConvers.pdf>, <http://www.wbur.org/commonhealth/2018/02/14/mit-brain-study>.

<sup>5</sup> <http://www.ascd.org/publications/books/61189156/chapters/The-Growing-Need-for-Interdisciplinary-Curriculum-Content.aspx>.

<sup>6</sup> <http://www.pz.harvard.edu/projects/multiple-intelligences>.

<sup>7</sup> [http://www.earlychildhoodnews.com/earlychildhood/article\\_view.aspx?ArticleID=607](http://www.earlychildhoodnews.com/earlychildhood/article_view.aspx?ArticleID=607).

Underlying the design of the curriculum are principles of backward design and those of the Universal Design for Learning framework, as well as paying particular attention to culturally sustaining practices. We are working on:

- writing
- programs that link school to home
- observational assessment
- dual language platforms
- overall coherence for pre-K to second grade, with a particular focus on English Language Arts standards

You can explore any of our curricular and other materials on our early childhood website: (<https://sites.google.com/bostonpublicschools.org/earlychildhood>.)

*Seventh, use NAEYC accreditation as a driver to set quality at the school level.* When I was at the Department of Education administering preschool grants, NAEYC accreditation was a requirement for programs to receive a grant. The notion was that a nationally recognized outside organization had a better chance of validating quality than the local community or the state government (e.g., via QRIS). When I accepted the job at the BPS, one of the first thresholds of quality I mentioned to the mayor and superintendent was accreditation; it helped that accreditation was supposedly a requirement to receive a \$2 million grant that added funds for a part-time paraprofessional in kindergarten classrooms. Although the requirement was not truly mandated, I used it as tool to underscore the importance of quality at the district level. This is a good example of how state policies can align to help improve programs.

In 2007 we started our accreditation work in earnest in 15 schools. We intentionally selected schools that ranged in size, that posed different challenges to procuring accreditation, and that had different levels of motivation with respect to earning accreditation. Initially we hired outside “mentors” who had worked with community-based programs, but we quickly learned that this was not our best strategy. We found that some of the mentors would do all of the work for the schools, not allowing them to swim on their own. We also found that too many of the mentors were treating the accreditation criteria as a checklist and not as reflective practice necessary to sustain change. We decided to change our partnership with outside mentors structurally in two ways: we partnered them with a BPS coach, and we held monthly meetings with the BPS coaches and mentors to calibrate the work. We also developed an NAEYC methodology that moved the work to a deeper and more reflective space than the checklist approach. It is important to keep in mind here that while piloting work in a district is a luxury that allows you to learn with schools, there can be drawbacks, as there is urgency to the work and the possibility of a change in course direction in leadership or funders.

The costs of NAEYC supports in Boston are not trivial. We spend around \$6,000 per classroom each year, and it usually takes 3 years to achieve accreditation. We now have over 40 accredited schools. To fund this work, we have used a combination of district and private money.

We are now at a crossroads with the NAEYC. Our early childhood programs go up to second grade, but the NAEYC is primarily focused on pre-K and kindergarten. As a department that is now responsible for 15,000 students, 70% of whom are on free and reduced lunch plans, we need a validation system to support all of our early childhood students. We are currently thinking through our options: maintain (but perhaps expand) the NAEYC system, adopt another K-12 accreditation system, or develop our own.

*Eighth, whether degrees are critical for education workers is a fraught issue.* A large number of early education workers lack bachelor's degrees, and less than a sliver have master's degrees. The work of educating and cultivating young learners is complex. Every day we ask teachers to emotionally support children, facilitate their conceptual knowledge, and crack the complex codes of reading, writing, and math. This work requires creativity, flexibility, observation, reflection, classroom management, planning, content knowledge, and an ability to respect and understand a variety of cultures that influence behavior and learning styles. Teaching is hard, and currently the data indicate that for pre-K to third grade we are not doing it well. National studies that have been conducted using the Classroom Assessment Scoring System place teachers somewhere in the 3s (on a scale of 1-7) on instructional supports, conceptual development, and language modeling.

Perhaps 20 years from now we will wonder how this work was ever done by anyone with less than a master's degree and a 2- to 4-year residency, but in today's reality the field is reluctant to require degrees and has no preservice placement requirement. The reluctance comes from the paucity of evidence around degrees, fear of losing diversity, and difficulty in finding qualified staff who are willing to work long hours for little pay. Also, people know intuitively that a degree does not make a teacher. Rather, it is in part a matter of personality traits, though it takes much more than personality; it also requires, for example, reflection, planning, and persistence.

That many early education teachers do not have degrees is also in my view connected to the fact that early education and care are often born from programs that are designed to help parents work and that are supported either through subsidies or by parent fees. Both sources of funding limit the ability to pay teachers and both pit access against quality. To be sure, the growth of the universal preschool movement is changing that, but progress is slow. To mitigate this problem, I believe preschool and its related educational requirements/certifications and compensation need to be included under the auspices of public education. This does not necessarily mean that preschool has to be delivered by the public schools: programs in New Jersey, Tulsa, New York City, and Boston offer some examples of successful mixed-delivery programs. Formally linking public schools and early education programs

**Formally linking public schools and early education programs will not only improve compensation, PD, and supports but will also provide many more opportunities to create meaningful linkages with birth to third grade programs and to transform public education from kindergarten to third grade.**

will not only improve compensation, PD, and supports but will also provide many more opportunities to create meaningful linkages with birth to third grade programs and to transform public education from kindergarten to third grade.

*Last, creating a pre-K model for community-based programs is crucial.* When the BPS opened up free preschool to 4-year-olds in the city it created an economic challenge to community-based preschool programs. (Preschool is the most economically sustainable due to large ratios). The BPS quickly became a

large part of the market, moving from serving around 10% of 4-year-olds to serving 55%. Teachers with BA degrees often applied for BPS jobs over community-based program jobs. Compounding the problem was that families who wanted a more “desirable” school had to apply to preschool (K1) in that system, as it increased their chance of getting their child enrolled in this school later on. This dramatic change was a disruptive influence and created tension between community-based organizations and the BPS. It also put families in the challenging position of having to choose between access, quality, and their child’s K-12 experience.

The new mayor is moving in the direction of expanding preschool programs in both the public schools and in community-based programs. To assure families of equity in quality, the mayor has designated a task force to oversee the design of a mixed-delivery system. We are excited about creating a “connective” system between community-based organizations and the BPS, as it would help programs develop meaningful pathways for students that would allow information to go from teacher to teacher and directors to principals, thereby improving overall communication to families. The opportunity for schools and community-based organizations to become more interdependent on one another is also exciting; for example, if a program is funded then families in community-based organizations would come off of the BPS waitlist. Finally, this might allow us to help support 0-3 programming, which is largely structurally ignored by the public school system.

I am often asked about the cost of public schools versus cost of community-based programs, as policymakers want to weigh cost and benefit and/or how much “quality” costs. The challenge of answering these questions is that the costs to the BPS and each city and town are relative to their context. The work in community-based programs, with coaching, BA-comparable salaries, and 12 months a year for 8 hours a day, costs the same per child as that in the BPS system, if not more. In any event, the current state and federal reimbursement rate is around 60% of that cost, so much more work will have to be done to combine (or braid) funds to cover the real price of investing in early childhood education. Our current universal pre-K budget is around \$11,000 per child for community-based organizations, with an additional \$7,000 coming from state subsidies to cover wraparound services and nonschool days. The universal pre-K program pays teachers BPS starting salaries and provides access to comprehensive services.

## CONCLUSION

---

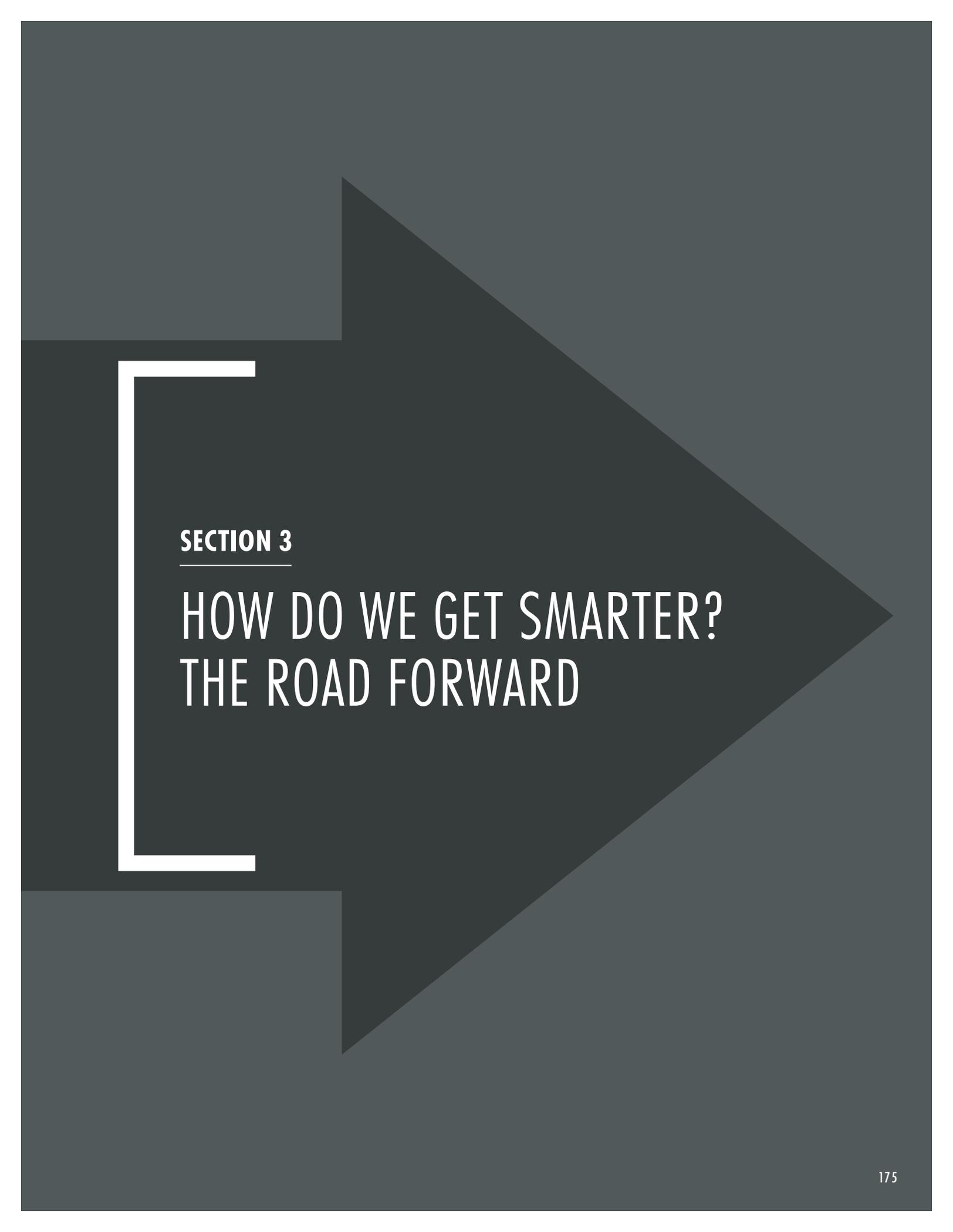
My motivation for writing this chapter is to help other programs think through the steps necessary for change, which include being systematic, collecting data, staying on task, and giving staff room to grow and solve problems. That said, our team will change course and revise our strategies, methods, and partners as needed. But we do so within a framework we created for ourselves that is centered on curriculum, professional development, coaching, and partnerships.

Finally, I would like to thank the leadership of the BPS for their support of the work. I would also like to give a large thank you to the staff of the Department of Early Childhood; we have a small, determined group of people, and the focus and passion they give to their jobs and ultimately to students is tremendous. They have an incredible wealth of knowledge and expertise, and day in and day out they show themselves to be stubborn, humble, and true leaders in the field.

## References

---

- Bardige, B., Baker, M., Mardell, B. (2018). *Children at the center: Transforming early childhood education in the Boston Public Schools*. Cambridge, MA: Harvard Education Press.
- Borman, G. D., & Dowling, N. M. (2006). Longitudinal achievement effects of multiyear summer school: Evidence from the Teach Baltimore randomized field trial. *Educational Evaluation and Policy Analysis, 28*(1), 25-48.
- Checkoway, A., Goodson, B., Grindal, T., & Hofer, K. (2016). *Year 1 Massachusetts Preschool Expansion Grant (PEG): Evaluation Report*. Cambridge, MA: Abt Associates.
- Entwisle, D. R., & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review, 57*(1), 72-84.
- Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics, 86*(1), 226-244.
- Jan, T. (April 7, 2007). Boston preschools falling far short of goals, study says. *Boston Globe*.
- Minnesota Department of Human Services (2005). *School readiness in child care settings: A developmental assessment of children in 22 accredited child care settings*. St. Paul, MN: Minnesota Department of Human Services, Children and Family Services.
- Sachs, J., & Weiland, C. (2010, September). Boston's rapid expansion of public school-based preschool: Promoting quality, lessons learned. *Young Children, 74*-77.
- Weiland, C., Ulvestad, K., Sachs, J. & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly, 28*(2), 199-209.
- Weiland, C., & Yoshikawa, H. (2013). The impacts of an urban public prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development, 84*(6), 2112-2130.
- Weiland, C., & Yoshikawa, H. (2014). Does peer socio-economic status predict children's gains in receptive vocabulary and executive function in prekindergarten? *Journal of Applied Developmental Psychology, 35*(5), 422-432.
- Whitebook, M., Sakai, L., Howes, C. (1997). *NAEYC accreditation as a strategy for improving child care quality: An assessment by the National Center for the Early Childhood Work Force*. Washington, DC: National Center for the Early Childhood Workforce.
- Yudron, M., Weiland, C., & Sachs, J. (2016). *BPS K1DS: Piloting the Boston Public Schools' prekindergarten model in community-based organizations*. Boston: Boston Public Schools. [http://bpsearlychildhood.weebly.com/uploads/1/0/1/3/10131776/bpsk1ds\\_final\\_report\\_feb2016\\_11.pdf](http://bpsearlychildhood.weebly.com/uploads/1/0/1/3/10131776/bpsk1ds_final_report_feb2016_11.pdf).



**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.<sup>1</sup> 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.

---

<sup>1</sup> 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](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).

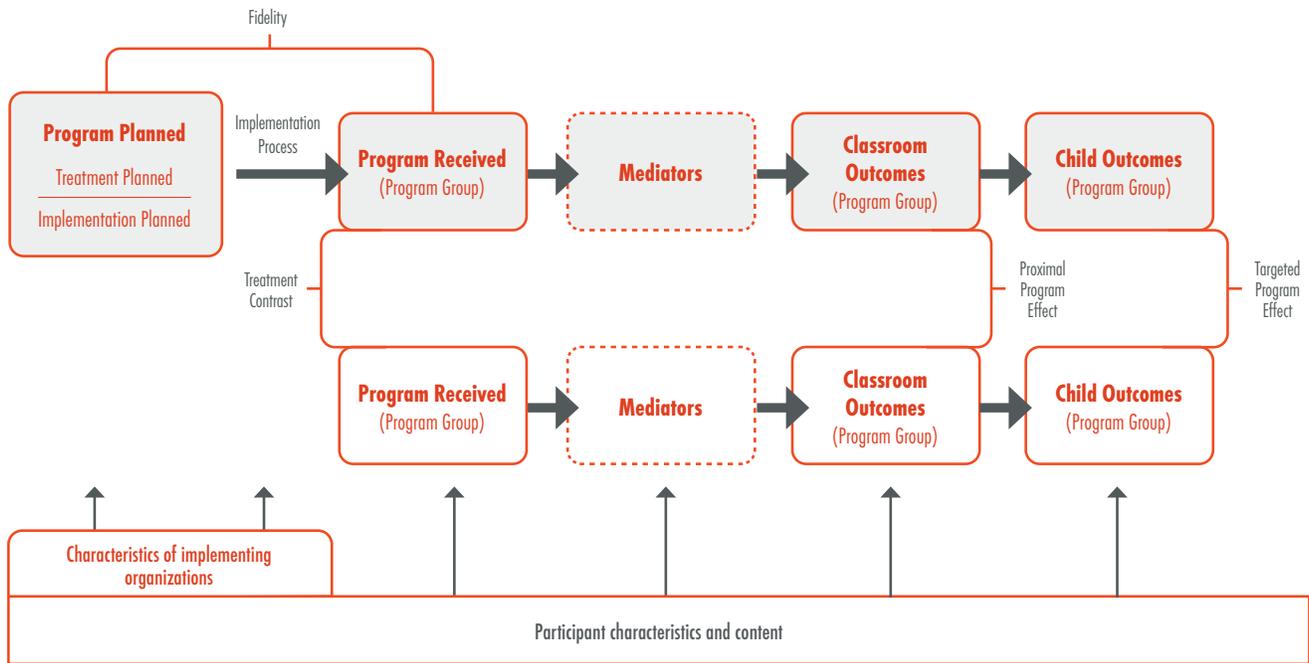
This chapter aims to help design strong implementation research to complement rigorous evaluation of ECE programming. It, therefore, has two goals: to provide a set of frameworks to help guide the empirical study of program implementation in an evidence-building context and to discuss potential methodological 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](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, & Koulish, 2014; Schaack, Tarrant, Boller, & Tout, 2012; Tout, Epstein, Soli, & Lowe, 2015; Wesley & Buysse, 2010; Young, 2017), there is growing interest in the burgeoning field of improvement science and its promise to promote a culture of quality improvement in early childhood settings (Boller, Sciarrino, & Waller, 2018; Daily et al., 2018; Hetzner, Arbour, Douglass, Mackrain, & Agosti, 2018). Like implementation science, improvement science has been used extensively in health care (Grol, Baker, & Moss, 2002; Improvement Science Research Network, 2010; Institute for Healthcare Improvement [IHI], 2003). Improvement science uses foundational concepts developed in business and manufacturing (Deming, 1986) and also draws on implementation science, systems theory, behavioral science, and change management (Daily et al., 2018). It has expanded to disciplines including education, child trauma, and child welfare (Agosti, Conradi, Halladay Goldman, & Langan, 2013; Bryk, 2015; Ebert, Amaya-Jackson, Markiewicz, Kisiel, & Fairbank, 2012; Haine-Schlagel, Brookman-Frazee, Janis, & Gordon, 2013). Although QI initiatives in ECE are growing more common, how such initiatives are defined and implemented varies widely across ECE settings and systems (Daily et al., 2018; Derrick-Mills et al., 2014). Few early childhood researchers or ECE practitioners interested in quality improvement

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

Because implementation science and improvement science are new to the early childhood field, researchers may be confused about what taking an implementation science or improvement science perspective means when studying the effectiveness, adaptation, and/or scale-up of early childhood programs, policies, or practices. Furthermore, policymakers, practitioners, and researchers may struggle to understand how a study focused on implementation

or quality improvement differs from what program evaluators have been doing for years when they study for whom and under what conditions early childhood programs and systems achieve their best results. In this chapter, I outline the commonalities and distinctions between implementation science and improvement science, and I demonstrate how they can enhance program development and program evaluation in early childhood settings. I contend that implementation science and improvement science, though distinct, share many common elements and are highly compatible. An understanding of what these different frameworks offer, in both their commonalities and

their unique features, can support effectiveness and continuous improvement of programs, policies, and practices (hereafter referred to collectively as "interventions") in the early childhood field.

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

## COMPARISON OF IMPLEMENTATION SCIENCE AND IMPROVEMENT SCIENCE

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

### ► Definitions and main aims

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

<sup>1</sup> 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).<sup>2</sup> 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.<sup>3</sup> 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

---

<sup>2</sup> 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).

<sup>3</sup> 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**

<b>Areas of focus</b>	<b>Implementation Science</b>	<b>Improvement Science</b>
Systematic study of practices to achieve improvements in outcomes	√	√
Local context	√	√
Real-world settings	√	√
Adaptation	√	√
Innovation	√	√
Intervention fidelity	√	
Implementation fidelity	√	√
<b>Aims</b>		
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.<sup>4</sup> 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.

---

<sup>4</sup> 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).

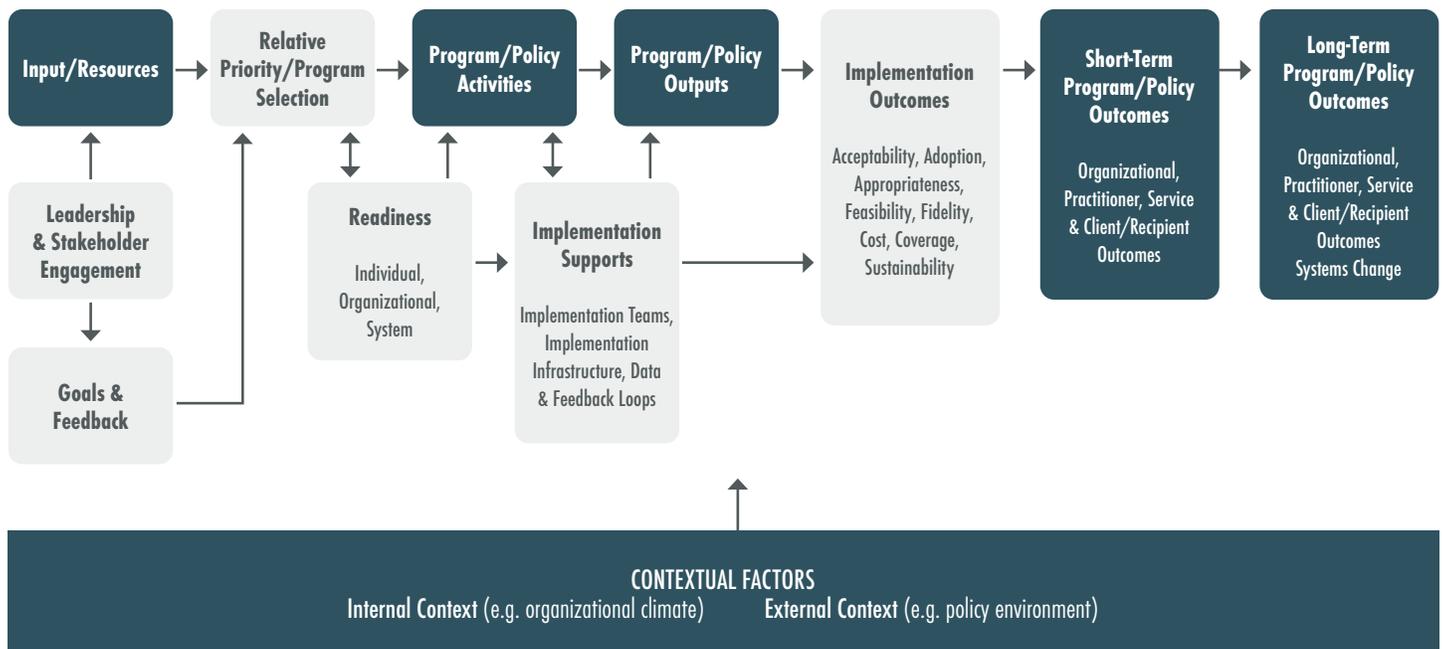
Another contribution that implementation science has made to traditional program evaluation is its treatment of implementation outcomes as distinct from intervention outcomes (Peters, Tran, & Adam, 2013; Proctor et al., 2011). Proctor and colleagues (2011) distinguished implementation outcomes from service 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).<sup>5</sup> 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.<sup>6</sup>

<sup>5</sup> 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).

<sup>6</sup> 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**

<b>Research Questions/ Outcomes of Interest</b>	<b>Implementation Science</b>	<b>Improvement Science</b>
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,<sup>7</sup> 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.

---

<sup>7</sup> 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.<sup>8</sup>

---

<sup>8</sup> 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.<sup>9</sup> 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.

---

<sup>9</sup> 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).<sup>10</sup>

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.<sup>11</sup> 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.<sup>12</sup> 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.

---

<sup>10</sup> 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).

<sup>11</sup> 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).

<sup>12</sup> 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.<sup>13</sup> 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.

---

<sup>13</sup> 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.<sup>14</sup> 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).

<sup>14</sup> 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.

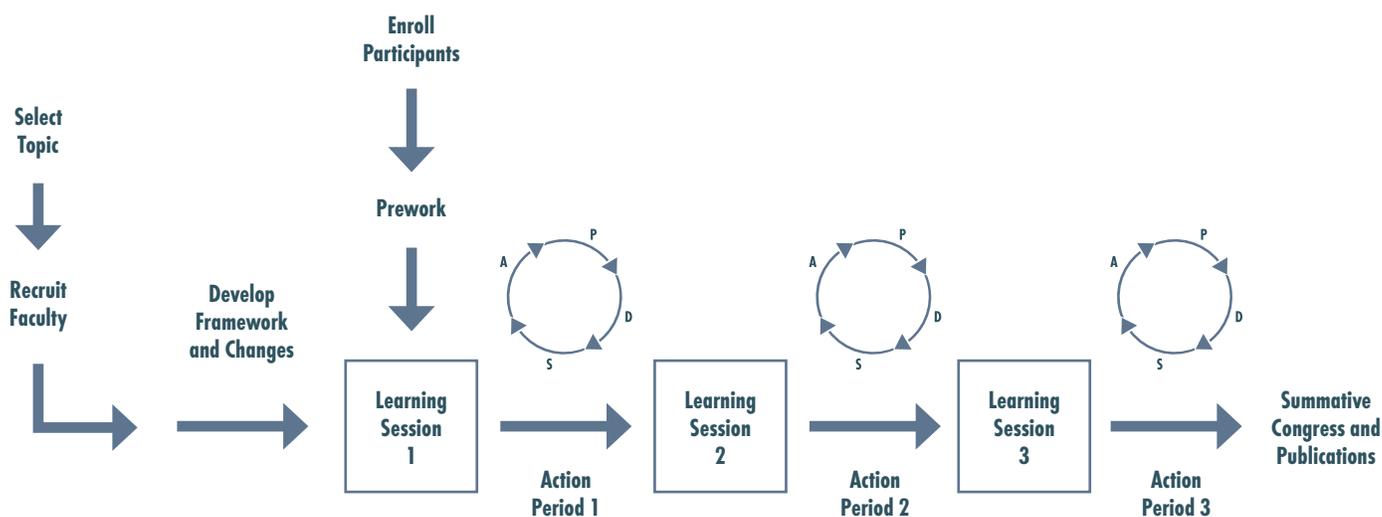
## ► Improvement science

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.<sup>15</sup>

Figure 2. Improvement science methodology.

### IHI Breakthrough Series Collaborative Model



Source: Reprinted from [www.IHI.org](http://www.IHI.org) with permission of the IHI, ©2018.

<sup>15</sup> 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.<sup>16</sup> 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.<sup>17</sup>

In total, the HV CoIIN engaged multidisciplinary teams from 13 MIECHV awardees<sup>18</sup> 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).

---

<sup>16</sup> 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.

<sup>17</sup> 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.

<sup>18</sup> 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.<sup>19</sup> 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.

---

<sup>19</sup> 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.<sup>20</sup> 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.

---

<sup>20</sup> For more information, see [http://hv-coiin.edc.org/sites/hv-coiin.edc.org/files/HV%20CoIIN%20Information%20Resource%202017\\_0.pdf](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](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](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](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](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](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](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](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](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](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](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.<sup>1</sup> 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).

---

<sup>1</sup> 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).<sup>2</sup> 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.**

---

<sup>2</sup> 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?

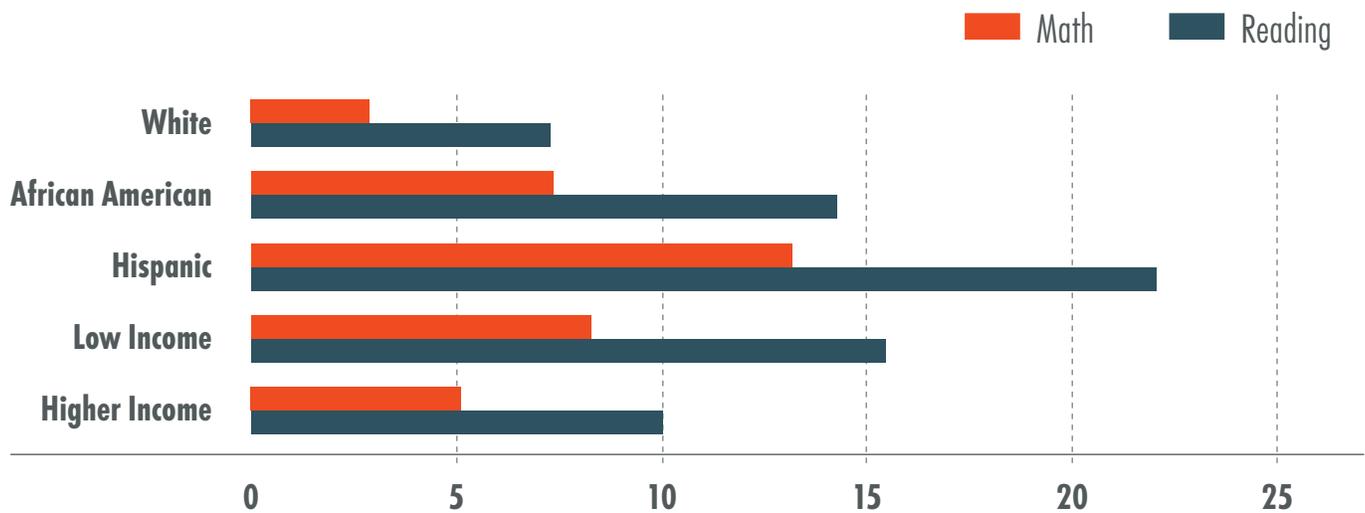
## EARLY CHILDHOOD PROGRAM EFFECTS AND EQUITY

---

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).<sup>3</sup> 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.

---

<sup>3</sup> 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](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
<b>Context</b>	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
<b>Perspective</b>	<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>		
<b>Program</b>	<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
<b>Design, Sample</b>	<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>
<b>Procedures</b>	<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><b>Analyses, Dissemination</b></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>



**REFLECTIONS AND INSIGHTS**

# MOVING ECE IMPLEMENTATION RESEARCH FORWARD

---

Sara Vecchiotti, Ph.D., Esq., Foundation for Child Development



Implementation research is applied research. A rigorous scientific approach must be used to take account of the complexities of implementing programs and policies, in real time, for specific populations under specific conditions. Implementation research thus brings opportunities and challenges. This volume does not prescribe a single definition of implementation research. Instead, it draws attention to the potential of implementation research designs to fully investigate early care and education (ECE) programs in context. It also outlines how implementation research can advance the field of ECE by answering questions that are relevant to policymakers, which can help them better understand issues of equity, and so ensure that ECE programs produce positive outcomes for all young children. This chapter highlights the volume's main insights about what researchers should know and be able to do when they apply implementation research to further build evidence for the ECE field. Further, it rethinks the perspective and expectations for applied researchers seeking to engage in implementation research with policymakers and practitioners, and it emphasizes that shared operational knowledge is important for fostering collaborative research. Finally, it also underscores the need for future implementation research to prioritize strengthening the ECE workforce.

## WHY IMPLEMENTATION RESEARCH IN ECE?

---

Many of the chapters in this volume (Burchinal & Farran, Ch.1; Farran, Ch.4; Brooks-Gunn & Lazzeroni, Ch.2; Iruka, Ch.3) summarize what we have learned from past research about how to advance high-quality ECE for young children. For example, researchers have asked questions about program quality and effectiveness, short- and long-term outcomes for children, which specific program and system characteristics are tied to particular child outcomes, and how programs or systems can be implemented at scale and still produce the same benefits as smaller, more targeted landmark studies (Gomby, Larner, Stevenson, Lewit, & Behrman, 1995; Ramey & Ramey, 1998; Reynolds, Mann, Miedel, & Smokowski, 1997). Many of these questions, shared by both researchers and policymakers, are still relevant today (Jones & Vecchiotti, 2020; Phillips et al., 2017).

At the same time, new and refined research questions are emerging as certain contextual factors suggest an urgent need to expand applied research, using it to produce a more nuanced understanding of how programs and policies are being implemented and how they affect specific subgroups of children differently. Such an approach would represent a shift from focusing solely on end results or outcomes to figuring out what in a program's execution has led to those outcomes and why and how. Given this orientation, implementation research has the potential to answer questions that policymakers and practitioners prioritize as they seek to continuously improve or strengthen the ECE policies and programs that they govern, manage, and provide. To fully capture how ECE programs and policies influence young children's development, we must pay attention to both outcome- and implementation-oriented research.

### ► A practical approach for ECE

Much of the foundational ECE research used randomized controlled trials (RCTs) to assess causation and program impacts. However, RCTs alone may not allow for an in-depth consideration of the context or conditions that affect implementation quality, and they take a long time to produce results (Brooks-Gunn & Lazzeroni, Ch. 2; Halle, Ch.10). Unlike the early days of ECE research, it is no longer easy to find a “clean” control group, because so many children are in some type of care. Thus, it is difficult to compare children with no preschool experience to those with such experience (Brooks-Gunn & Lazzeroni, Ch. 2). Furthermore, the majority of state prekindergarten programs are implemented in mixed-delivery systems that encompass both public schools and community-based settings (Barnett et al., 2016). In today’s context, implementation research may be more practical than RCTs. It examines program implementation in real time while considering contexts and other variables that influence quality and outcomes, and it gives stakeholders and policymakers more timely answers (Halle, Ch.10).

Halle, Hsueh, and Maier (Ch.8) suggest that implementation research can also help achieve two ECE goals: scaling up effective ECE programs and ensuring better outcomes for all children. Evidence of program effects alone is not enough to successfully strengthen, replicate, scale, and sustain ECE programs and to meet the diverse needs of all children (Halle, Hsueh, & Maier, Ch. 8). Achieving such ECE goals necessitates “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” (p. 179). Implementation research goes beyond answering the question of whether effects are demonstrated to explaining why or why not.

### ► A bridge to understanding outcomes

ECE programs and policies are increasingly being brought to scale, particularly in states and municipalities (Friedman-Krauss et al., 2019). In the long term, if we cannot answer implementation scale-up questions related to how and when ECE is effective, we risk losing support for increased investment in ECE (Jones & Vecchiotti, 2020), because expectations for ECE to attain certain child outcomes might outstrip results (Brooks-Gunn & Lazzeroni, Ch. 2). Implementation research can help minimize this risk. As Maier and Hsueh assert in their contribution, strong implementation research is the key to achieving the positive child outcomes we see in small-scale model ECE programs when we turn to large-scale adaptations across populations and settings. Not only does implementation research ask what is happening—whether execution of a program or policy is accomplishing the stated purpose—it also asks how, why, and for whom a policy, program, or practice does or does not work (Maier & Hsueh, Ch.9). We urgently need implementation research to guide localities on the specific challenges and opportunities they may encounter as ECE programs are implemented in diverse real-world settings (Weiland, 2018). An implementation approach can also push ECE research forward by identifying deeper questions about the root causes of inequity and ways to eliminate disparities.

## ► Addressing inequity

As Iruka notes in her chapter, if we are to address persistent opportunity and achievement gaps, the field should cease “gap gazing” and blaming children as sources of disparities. Instead, we should investigate the root causes of such disparities and how ECE research-based practices and policies can help eliminate them. Nores adds that implementation research can measure the degree to which ECE programs and processes diminish or intensify inequities and unearth how program design or its implementation contributes to either result. Future investigations should ensure that “research components capture whether a program is working towards reducing inequities” and that those components are “validly defining these inequities in relation to the context and populations at hand”; they should also check that evaluations “are not introducing biases that reduce the chances of understanding whether the program works and, if it does, for whom” (Nores, Ch. 12, p. 279). Children of color will become the majority of children in the near future, the proportion of dual language learners will increase, and income-inequality will likely continue to grow, and so we need to accelerate implementation research that addresses equity issues and concerns in ECE as a way of promoting all children’s healthy development.

## FORGING AHEAD IN ECE IMPLEMENTATION RESEARCH: REFRESHING THE APPLIED RESEARCHER PERSPECTIVE

---

Conducting applied research has always been challenging. Yet implementation research in ECE is potentially even more difficult and multifaceted. As this volume demonstrates, conducting sound, rigorous, high-quality ECE implementation research to build evidence for the field is no easy task. Realistically, researchers doing such work need to be willing to “embrace the messy” from initial design through final analysis and interpretation. The messiness reflects the complexities of the interventions and is precisely what makes the work so interesting and signals that the issues and questions explored are not easily answered.

What perspective is likely needed to embrace the messiness in implementation research? First, researchers engaged in such work will need to have a deep appreciation for ever-evolving contexts, typically encompassing multiple layers of policy and programmatic decisions and surrounding conditions. Second, in order to answer nuanced and interrelated questions nested within and across contexts, researchers will likely need extensive knowledge about complex, rigorous designs and methods of analysis. Third, given the nature of implementation research, in order to produce findings and implications that are useful and meaningful researchers will also need to consider developing more collaborative relationships with research partners, policymakers, and practitioners. Without such relevance and responsiveness to policy and practice, research will be unlikely to provide useful evidence that can be used to improve high-quality early learning opportunities that meet the needs of the children served.

### ► Evidence-building in context

Implementation research seeks to gain knowledge not so much about what was done as about how it was done; it takes an evolutionary approach, considering past and present contexts relating to changing resource inputs and altering outputs as programs and policies are revised in real time, and it examines adaptation as programs and policies evolve in response to environmental contexts and conditions (Pressman & Wildavsky, 1984). It requires flexible, responsive approaches and so cannot rely on the clear-cut, stable approaches that guide causal impact analyses. Studying implementation is worthwhile particularly because, as Pressman and Wildavsky (1984) note, implementation “is a struggle over the realization of ideas”; there is, they add, “no escape from implementation and its attendant responsibilities” (p.180). ECE implementation research particularly emphasizes understanding context in detail—how context influences program implementation and how the interrelationships between implementation context and program model lead to variation in program effects and outcomes (see Sachs, Ch. 7, for a case study).

Implementation research is further distinct in its organization of scientific inquiry in two ways: with an inward focus that considers “a program’s theory of change or implementation processes” and an outward focus that attends to the “larger context and infrastructure supports that surround a program” (Hsueh, Halle, & Maier, Ch. 8, p. 182). This dual focus allows researchers to examine sources of variation that may contribute to program effectiveness and to child outcomes, including among subgroups of children (Hsueh, Halle, & Maier, Ch. 8). Implementation research questions are of particular interest to ECE policymakers, because the approach builds evidence for program effectiveness through a continuous iterative cycle of execution as the program model evolves, adaptation as the program model and system supports are refined, and evaluation as the program model is tested (Maier & Hsueh, Ch. 9). Thus, not only can results and feedback be provided in a timely manner, but we can also examine implementation across a range of ECE settings, contexts, and populations (Ryan, Ch.11; Hsueh, Halle, & Maier, Ch.8).

Such evidence-building research on ECE programs and policies within a specific context contributes to our growing knowledge of what works or not, for whom, and under what conditions. Many ECE implementation research efforts take place at the municipal or state level. Such a local/state focus aligns with the fact that many ECE programs and initiatives are locally designed and regulated. Ryan (Ch.11) stresses that acquiring information about the factors that contribute to successful programs also includes understanding how local conditions, and how program adaptations made by local leaders and actors, shape implementation and program improvement strategies.

Undertaking various qualitative case studies of state and local implementation can show the how of ECE programming in various communities, thereby helping to identify the factors that influence differences in implementation (Ryan, Ch. 11). Ryan also explains how rigorous qualitative implementation studies provide practical, in-depth contextual information about local culture, conditions, and factors that help elucidate why programs fulfill their promise or not.

Similarly, Hsueh, Halle, & Maier suggest that examining local resource variation in implementation can help us figure out how to strengthen, replicate, scale and sustain ECE programs. Findings from local or state implementation studies have implications for areas with similar characteristics (Nores, Ch. 12). Empirical evidence from implementation case studies suggests that systematic relationships can emerge between different policy and program characteristics and among the problems encountered (Pressman & Wildavsky, 1984). Eventually, as knowledge about implementation in local contexts increases and similarities emerge, findings will serve to enhance ECE systems and programs nationally as well.

When conducting ECE implementation research in context, therefore, applied researchers will likely need to be comfortable with change, with challenge, and with responding to real-time circumstances that complicate the delicate balance between the inward and outward that a dual focus requires. It is not easy to develop or unearth linkages between theory and practice, and it can be even more challenging to assess such linkages in real-time, on-the-ground interventions. Further, researchers will need to be open to seeking knowledge about local program/policy context and to understanding how the nuances of that context influence their research inquires and design. It is not enough to know about programs or policies in general; researchers need to know how policies and programs are implemented or adapted locally.

### ► Rigorous and complex research design

Another area where researchers can “embrace the mess” is in research design. Accounting for implementation context calls for rigorous and complex mixed-methods research designs that respond to varying program scopes and scales within changing political landscapes—and researchers will likely need to know how to design and conduct such comprehensive, inter-related inquiries. As Halle points out, implementation research studies can be embedded in RCTs or can take the form of separate mixed-methods, quasi-experimental, or “innovative” designs (e.g., effectiveness implementation). Further, implementation research designs often use both quantitative and qualitative data sources so that they can fully describe and examine the constructs of interest, the relationships among constructs, similar and differential impacts on subgroups, the influencing and mediating factors in execution, and individual perceptions, attitudes and experiences—all in unfolding stages of implementation and in changing contexts (Halle, Ch. 10; Hsueh & Maier, Ch. 9; Ryan, Ch. 11). As a result, the applied implementation research approach is not easy to design or carry out. It is complex work, but that challenge is also what makes such work interesting and relevant to the field. It has great potential to drive and support the continuous quality improvement of ECE programs and policies.

Implementation research must be “embedded” in existing program and policy activities for it to best examine context and therefore be effective (Halle, Ch. 10; Sachs, Ch. 7), and this adds another layer of complexity. Often, implementation research aims to support continuous quality improvement efforts to make ECE programs and

policies more effective (Halle, Ch. 10; Maier & Hsueh, Ch. 9; Sachs, Ch. 7). Useful and meaningful implementation research requires applied researchers to actively and constructively work throughout the entire research process with the decision makers—policymakers and/or practitioners—who are implementing and supporting the program or policy and are responsible for how the program processes work and for outcomes (Halle, Ch. 10; Sachs, Ch. 7). In this volume, Jason Sachs explains his story of building and scaling the Boston Public Schools' prekindergarten-2nd grade program by intentionally using research to inform the process of change. His narrative reflects the realities encountered while conducting implementation research; researchers must “think through the steps necessary for change, which include being systematic, collecting data, staying on task, and providing staff room to grow and solve problems. That said,” he continues, “our team will change course and revise our strategies, methods, and partners as needed. But we do so within a framework we created for ourselves that is centered on curriculum, professional development, coaching, and partnerships” (p. 173). Across all its stages, implementation research requires a collaborative relationship between those examining the program/policy and the stakeholders supporting and applying the program or policy in practice in real-time (Pressman & Wildavsky, 1984). Given the nature of such research and the nature of kind of research that the ECE field is more used to doing, we may need to refresh our perspective on what is required to establish and maintain collaboration.

## **COLLABORATION: ESSENTIALS OF WORKING RESEARCHER-POLICYMAKER RELATIONSHIPS**

---

As the ECE field continues to consider how best to establish, improve, and scale ECE programs and systems, policymakers and researchers are joining forces to determine what works, for whom, and under what conditions. In such work, researchers will likely need to acknowledge that many policymakers and practitioners operate within a high-pressure environment of just “keeping things going” (Pressman & Wildavsky, 1984, p. 172). ECE researchers and their partners also face challenges from financial constraints (National Academies of Sciences, Engineering, and Medicine, 2018) that ECE programs and policies are subject to, and sometimes they also face difficulties from heightened political attention and public scrutiny (Bardige, Baker, & Mardell, 2018). As a result, researchers and policymakers are operating in a realm where expectations for program and child outcomes must be managed (Brooks-Gunn, Ch. 2), especially given limitations in program investment, infrastructure support, service model comprehensiveness, and duration. As such applied researchers will likely need a deep commitment to building and maintaining research collaborations in such conditions before, during, and after their research investigations.

Whether implementation research is conducted through a research-practice partnership (RPP) model<sup>1</sup> or through another formal arrangement, applied implementation research must be collaborative in nature, especially when it aims to continuously build, refine, and scale ECE programs and policies in practice (Sachs, Ch. 7; Hsueh & Maier, Ch. 9). Research studies from the Institute of Education Sciences' Early Learning Network<sup>2</sup> (ELN) exemplify a collaborative structure in which policymakers, practitioners, and researchers are partnering to examine ECE issues relevant to implementation in individual studies and as a network. The ELN also supports research conducted through a long-standing RPP among the Boston Public Schools, the University of Michigan, MDRC, and the Harvard Graduate School of Education, profiled by Hsueh and Maier in this volume. New York City's Early Childhood Research Network, made up of representatives from multiple city agencies and researchers from several institutions of higher education, was also built to answer codesigned questions relevant to the ECE workforce in the context of scaling up a full-day universal prekindergarten program (Foundation for Child Development, 2018; Hsueh & Maier, Ch. 9).

To generate useful evidence that can improve practice, research questions must be highly relevant to the topics that interest the stakeholders who are responsible for supporting and implementing a program or policy (Tseng & Nutley, 2014; Tseng, Easton, & Supplee, 2017). At times, this may mean that the question being studied is not aligned with the question of most interest to the researchers; researchers must therefore be flexible and responsive if they are to provide data that can guide policy decisions. In such research, clear roles and interests are traditionally defined for the researcher, the policymaker (elected official, political appointee, career staff), and practitioners based on their separate domains of expertise and responsibility that help to navigate the researcher-stakeholder relationship (Zervigon-Hakes, 1995). Though it is true that researchers, policymakers, and practitioners are from "different worlds" (Zervigon-Hakes, 1995), the work they are joining forces to do requires intersecting knowledge across many areas of expertise, meaning that such collaborations can be challenging to navigate. Collaboration rests on a grounded or shared understanding between researchers and stakeholders about the research purpose, design, and course of work. For a collaboration to be productive or successful each party has the responsibility to acquire operational working knowledge of the context in which the partners live. Recognizing and respecting each party's expertise in an RPP is critical to successful collaboration (Henrick et al., 2017). Yet such recognition and respect is a minimum threshold; shared operational knowledge further extends the notion of effective collaboration.

---

<sup>1</sup> See Henrick, Cobb, Penuel, Jackson, & Clark (2017) for description of various RPP models.

<sup>2</sup> The goal of the Early Learning Network is to "positively impact the lives of children in preschool through Grade 3 by investigating the implementation of early learning policies and programs; identifying malleable factors associated with early achievement; and providing information, tools and products that policymakers and practitioners can use to build effective early learning systems and programs" (<http://earlylearningnetwork.unl.edu>).

## ► Shared operational knowledge

Shared operational knowledge builds on the idea that collaborative research is a two-way street (Tseng et al., 2017). Such mutual understanding can inform and refine the research process at every stage, from question identification, planning and protocol development, through exploration and the interpretation of findings and considering policy implications, thereby increasing the translation of knowledge across fields and the likelihood that research findings will change practice and policy. Building a collaborative process that encompasses the entire research process rather than just certain stages of the process is an important shift in approach for applied researchers. By providing a common ground, a shared understanding can also help to build and maintain trust

**Building a collaborative process that encompasses the entire research process rather than just certain stages of the process is an important shift in approach for applied researchers.**

among collaborators through research study design, data collection and analysis, and interpretation of findings. Shared understanding can also increase the clarity of the communications—discussing the work itself and the implications of the findings—that are essential to supporting healthy collaboration. It can also help to manage appropriate expectations—especially in the understanding of applied researchers—regarding what research can do to influence or support continuous quality improvement of ECE programs and policies and how it can do it. Research findings can be used in policy and

program decision making in various formal and informal ways. Applied researchers also need to understand how internal and external contextual factors influence how feasible it is to adopt suggested policy and program changes, the timing of their adoption, and the capacity for adopting them in both the near future and the long run.

As different forms of collaboration likely involve different types and levels of shared operational knowledge, implementation researchers need to be flexible and willing to change parameters. The necessary type and level of shared operational knowledge will vary for each scientific inquiry and for each program or policy under study. Crucially, operational knowledge means a working, functional understanding and familiarity; it does not mean having either just a rudimentary knowledge or deep expertise about other researchers' and stakeholders' content domains.

In part, applied implementation research is about developing the capacity to learn how programs and policies are executed in practice (Pressman & Wildavsky, 1984); thus, much operational knowledge is about practice. Researchers who engage in this work need to consider building their own knowledge about the fundamentals of partners' and stakeholders' work, especially because they are already examining the tensions between the planned ideal and actual implementation. For example, if they are to make meaningful policy recommendations, researchers need to understand a program or policy's specific purpose, elements, and processes. Some examples of shared operational knowledge that represent elements that impact the decision making of stakeholders include:

- the political and funding environments (whether the program and political leadership is invested in the program's success or failure, other priorities of the administrative leadership that may exert an influence, whether program funding reflects the true cost of the quality of care, whether the system is operating at a fiscal deficit or in lean ways, what issues or concerns local ECE advocates are addressing, the level of scrutiny in the local news media);
- the governing rules and regulations (potentially across several agencies involved in fragmented ECE systems and involve the policy or program service goal such as access or quality, or involve eligibility, enrollment, and attendance policies and procedures); and
- the program management aspects (why particular workforce or parent supports are in place or not, reasons for particular staffing models, caseloads or teacher-child ratios, why particular curricula or assessments are in place, characteristics of program staff and those served by the program, whether different populations of children have different access to services and why, past program initiatives that succeeded or failed).

All such contextual and operational considerations help shape a program and drive its implementation. Lacking such shared operational knowledge, researchers may find it difficult to collaborate with stakeholders to conduct program-scaling implementation studies addressing the policy and practice research questions.

A shared understanding can also help researchers maintain and nurture collaboration and trust needed among partners through all the stages of applied implementation research. Just as implementation comprises the stages of exploration, installation, initial and full implementation, the research study itself has stages that require continued collaboration. Collaboration with research partners does not end with the co-construction of research questions but continues through the finalizing of research designs, cooperation in data access and collection, and data reporting. In formal RPPs, collaboration goes further and includes joint interpretation of data findings and the co-development of policy and practice recommendations that are suggested by the research findings. Researchers may need to adopt a new perspective—they may need to step back from being “expert” researchers providing one-sided recommendations for program and policy changes to stakeholders and instead come to see themselves and their policy and program partners as drawing on their distinct and shared expertise to contemplate the research findings and form ideas for continuous quality improvement together. Jointly interpreting the data and determining the implications of the findings also helps guide collaborative thinking about how to account for particular implementation contexts and can provide more insights into research-to-practice connections.

To promote open, productive dialogue in these joint deliberations—and to fully realize the potential of research to shape policy and practice—trust and candidness between partners is required. Researchers need to communicate honestly but tactfully, especially when findings are challenging or unfavorable to the partner’s efforts, reputation, or political stake. Shared operational knowledge can help because it brings greater understanding of the issues, challenges, and stakes in play for the policymaker partners. Such shared operational knowledge may also help researchers manage any predetermined notions or advocacy agenda they have about the work at hand and the partner’s performance and capacity. A tactful communications approach utilized by the researchers and the stakeholders alike can lead to more long-term, honest, candid partnerships between researchers and policymakers. Thus, researchers who engage in ECE implementation research—which is deeply embedded in local context and collaborative partnerships—may need a fresh research perspective. Applied researchers may need to learn to appreciate context and complex research designs, embrace struggle and change, welcome new areas of knowledge outside their comfort zone, and employ new joint working relationships and diplomatic communication methods to establish, sustain, and nurture collaboration with partners such as policymakers. Armed with such knowledge, skills, and dispositions, applied implementation researchers can increase the potential of research to shape, improve, or transform ECE policy and programs in ways that allow these programs to better serve children and their families.

## **FORGING AHEAD IN THE REAL WORLD: A RESEARCH AGENDA FOR ECE IMPLEMENTATION RESEARCH**

---

Policy- and practice-relevant implementation research questions related to the preparation, well-being, compensation, and ongoing professional learning of the ECE workforce are essential to continuous quality improvement in ECE programs and policies. Moreover, this area is primed for future implementation research, and the Foundation for Child Development is emphasizing it as a priority. The Foundation defines the ECE workforce as the professionals who educate and care for young children across a variety of settings (center and home-based) and systems (regulated and informal), as well as the individuals who provide leadership and support to them (e.g., lead teachers, coaches, home visitors, and administrators). The ECE workforce plays a significant role in the lives of young children in ECE programs, since the quality of their interactions with those they serve and the environmental stimulation that they provide directly influences children’s learning and development. Strengthening the ECE workforce will not only enhance the quality of early learning experiences, but lead to stronger outcomes for young children to help them meet their developmental potential. The Foundation’s ECE implementation research agenda centers on achieving the following goals:

- professionalize the early childhood field and build greater awareness of the status of the ECE workforce,
- enhance the quality of professional practice, and
- improve early educator preparation and ongoing professional learning.

Examples of implementation research questions related to the Foundation's ECE workforce goals that can generate empirical evidence of interest to the Foundation include questions found in the Young Scholar Program guidelines.<sup>3</sup> In addition, many of the recommendations made by authors in this volume for future research have implications for the ECE workforce and align with the Foundation's questions. Burchinal and Farran suggest that the field should move beyond assessing process quality elements when it examines program effectiveness and instead explore specific evidence-based instructional practices and evidence-based curricula content and how they relate to children's development. Such a focus on instruction can help increase the knowledge and skills of the ECE workforce.

Nores adds that what occurs in classrooms (e.g., practice, interactions, curricula content) 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," she adds, "are present as early as preschool and kindergarten, whether it be in teachers' perceptions . . . or children's own perceptions" (p. 278). Therefore, research could measure the degree to which ECE program design and the classroom practices of the ECE workforce diminish or perpetuate inequities. Moreover, if we are achieving educational equity and providing high-quality ECE for dual language learners, research must define the appropriate knowledge and competencies for ECE professionals who work with these children. We particularly need to understand how to implement effective program language models, instructional practices, and continuous assessment practices (Espinosa, Ch. 6). Pianta and Hamre suggest that we need more research on how to scale effective professional development systems. Specifically, research should explore the focus and purpose of professional development in relation to specific practice outcomes, the specific supports, intensity, and duration needed to enhance classroom instruction, and the effectiveness of course-based professional development and using certified providers.

Given both the diversity of the ECE workforce (Whitebook, McLean, Austin, & Edwards, 2018) and the children served and the fact that there are few people of color in ECE leadership positions, Iruka argues that additional research should explore access and supports for leadership opportunities in ECE programs, schools, and systems. Such research could tell us how to strengthen programs and schools by including diverse perspectives, how to create environmental climates valuing people of color, and how to promote equitable upward mobility. Many of the authors advocate continuing research to explore the impact of ECE workforce inequities in terms of compensation, work environments and benefits, and professional support—especially in relation to teacher well-being, turnover, and retention. By following these directions, implementation research could guide us in strengthening and better supporting the ECE workforce in their work with young children.

---

<sup>3</sup> For the Young Scholar Program guidelines please see: <https://www.fcd-us.org/about-us/young-scholars-program/>

## CONCLUSION

---

Implementation research is complex and rigorous in its questions, design, and methods as it seeks to untangle how context influences program and policy execution and the intended outcomes. Applied implementation research in ECE it is not easy, and researchers must embrace the messiness involved in such conceptually complex work. Yet such messy investigations help answer the field's questions about how to ensure that high-quality ECE programs promoting young children's development can become the norm and not isolated exemplars. By committing to meaningfully explore the how in studies of what works (or not), for whom, and under what conditions, ECE can better serve all young children in all settings. Such a commitment also likely entails building collaborative relationships with policymakers and practitioners—that is, the decision makers and implementers who are responsible for and can change ECE policy and practice. To nurture this collaboration, researchers will need to build their own understanding of the operational knowledge that is key to the experience of the policymakers and practitioners. If shared knowledge informs research questions, design, methods, data collection and analysis, interpretation of findings, and discussions of implications, the result will be more useful and effective studies that can change policy and practice. This work may not be for the faint of heart. However, engaging in real policy and practice problem solving is one way that researchers can work to ensure that young children experience high-quality ECE programs that help them meet their full developmental potential.

## References

---

- Bardige, B., Baker, M., & Mardell, B. (2018). *Children at the center: Transforming early childhood education in the Boston Public Schools, Cambridge, MA*: Harvard Education Press.
- Barnett, W. S., Friedman-Krauss, A. H., Gomez, R. E., Horowitz, M., Weisenfeld, G. G., Clarke Brown, K. & Squires, J. H. (2016). *The state of preschool 2015: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research. <http://nieer.org/state-preschool-yearbooks/the-state-of-preschool-2015>
- Foundation for Child Development. (2018). *The New York City Early Childhood Research Network: A model for integrating research policy, and practice*. New York, NY: Author.
- Friedman-Krauss, A. H., Barnett, W. S., Garver, K., Hodges, K., Weisenfeld, G. G., & DiCrecchio, N. (2019). *The state of preschool 2018: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research. <http://nieer.org/state-preschool-yearbooks/2018-2>
- Gomby, D. S., Larner, M. B., Stevenson, C. S., Lewit, E. M., & Behrman, R. E. (1995). Long-term outcomes of early childhood programs: Analysis and recommendations. *The Future of Children*, 5(3), 6-24.
- Henrick, E. C., Cobb, P., Penuel, W. R., Jackson, K., & Clark, T. (2017). *Assessing research-practice partnerships: Five dimensions of effectiveness*. New York: William T. Grant Foundation.
- Jones, J. & Vecchiotti, S. (2020). Policy and political influences on early childhood education. In D. Gullo & E. Graue (Eds.), *Scientific influences on early childhood education: From diverse perspectives to common practices* (pp. 37-54). New York, NY: Routledge.
- National Academies of Sciences, Engineering, and Medicine (2018). *Transforming the financing of early care and education*. Washington, DC: National Academies Press.
- Phillips, D., Lipsey, M., Dodge, K., Haskins, R., Bassok, D., Burchinal, M., . . . Weiland, C. (2017). Puzzling it out: The current state of scientific knowledge on pre-K effects: A consensus statement. In K. Dodge (Ed.), *The current state of scientific knowledge on pre-kindergarten effects* (pp. 19-30). Washington, DC: The Brookings Institution.
- Pressman, J., & Wildavsky, A. (1984). *Implementation: How great expectations in Washington are dashed in Oakland*. Berkeley: University of California Press.
- Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist*, 53(2), 109-120.
- Reynolds, A. J., Mann, E., Miedel, W., & Smokowski, P. (1997). The state of early childhood intervention: Effectiveness, myths and realities, new directions. *Focus*, 19(1), 5-11.
- Tseng, V., Easton, J., & Supplee, L. (2017). Research-practice partnerships: Building two-way streets of engagement, *Social Policy Report*, 30(4), 1-17.
- Tseng, V., & Nutley, S. (2014). Building the infrastructure to improve the use and usefulness of research in education. In K. S. Finnigan and A. J. Daly (Eds.), *Using Research Evidence in Education: From Schoolhouse Door to Capitol Hill* (pp. 163-175). Switzerland, Springer.
- Weiland, C. (2018). Pivoting to the “how”: Moving preschool policy, practice, and research forward. *Early Childhood Research Quarterly*, 45(4), 188-192.
- Whitebook, M., McLean, C., Austin, L. J. E., & Edwards, B. (2018). *Early Childhood Workforce Index—2018*. Berkeley: Center for the Study of Child Care Employment, University of California, Berkeley. <http://csce.berkeley.edu/topic/early-childhood-workforce-index/2018>.
- Zervigon-Hakes, A. (1995). Translating research findings into large-scale public programs and policy. *The Future of Children*, 5(3), 175-191.

FOUNDATION  
FOR CHILD  
DEVELOPMENT

Connecting  
research, policy  
and practice to  
create systems  
that work.

[www.fcd-us.org](http://www.fcd-us.org)

475 Riverside Drive, Suite 248, New York, NY 10115 • T 212-867-5777

### Suggested citation

---

Foundation for Child Development. (2020). *Getting it Right: Using Implementation Research to Improve Outcomes in Early Care and Education*. New York, NY. doi: [https://www.fcd-us.org/assets/2020/06/GettingitRight\\_UsingImplementationResearchtoImproveOutcomesinECE\\_2020.pdf](https://www.fcd-us.org/assets/2020/06/GettingitRight_UsingImplementationResearchtoImproveOutcomesinECE_2020.pdf)