

Failed to Scale: Embracing the Challenge of Scaling in Early Childhood

Snigdha Gupta¹, Lauren H. Supplee², Dana Suskind³, and John List⁴

¹TMW Center for Early Learning + Public Health, University of Chicago

² Child Trends

³ Department of Surgery, TMW Center for Early Learning + Public Health, University of Chicago

⁴ Kenneth C. Griffin Department of Economics, University of Chicago; NBER

Science of Scaling and Early Childhood: Two of the Greatest Public Policy Issues of Our Time

In recent years, citizens and lawmakers have become increasingly enthusiastic about adopting evidence-based policies and programs. Social scientists have delivered evidence of countless interventions that positively impact people's lives. And yet, most programs, when expanded, have not delivered the dramatic societal impacts promised.

This is especially true in the field of early childhood. Research suggests that high-quality early childhood programs (those serving children from birth through age five) have the potential to reduce the inequities that plague our nation and rob too many children of opportunity—and that they can deliver a 13% return on investment in the process. Despite the research and evidence, however, few programs have followed through and delivered on this promise. In order to reap the individual and economic benefits of early childhood programs, researchers and practitioners must figure out how to take these programs from small-scale experiments and implement them at scale in a way that enables population-level impacts.

This book—at its heart an exploration of threats to and facilitators of scaling—uses the lens of early childhood to examine the topic for two key reasons: a) early childhood is a critical period that lays the foundation for an individual's life course, a period in which successful interventions have the power to allow children to reach their full potential and b) early childhood programs are implemented through a complex and notoriously fragmented system that makes it particularly challenging to study, implement, and scale programs in the field. We've chosen early childhood in hopes that focusing on disentangling challenges in a particularly vexing domain will lead to conclusions that resonate in all public policy domains that experience similar scaling challenges, including health, education, energy, climate, and employment.

This book will examine some of the most critical questions of our time related to scaling evidence-based programs: Why have we failed to bring evidence-based programs found to be effective in small-scale experiments to the population level? How can we scale such programs effectively? Perhaps more importantly, we hope the book stimulates dialogue between researchers, practitioners, funders, and policy makers, paving the path toward stronger partnerships that bring fundamental changes in research and program design. Ultimately, we hope it elevates the burgeoning field of evidence-based policymaking.

Why is Early Childhood So Important?

The first few years of a child's life are critically important and provide a unique window of opportunity to positively impact their physical, emotional, and cognitive growth. During these years, a child's brain develops rapidly—at the rate of 1 million neural connections per second, a rate never matched in later life. Healthy brain development during this period provides children with the foundation necessary for developing to their full potential, setting them up for success in school and the workforce, and allowing them to lead a healthy and stable life.

Research has shown that access to nutrition, strong and nurturing relationships with caregivers, and exposure to environmental stress and adversity have profound impacts on a child's early brain development. Tragically, millions of children around the world may lack access to nutrition, the nurturing relationships needed for healthy brain development, or both. Lack of nutrition and early stimulation has tremendous negative consequences on a child's physical, cognitive, and socio-emotional development, as does exposure to stress. These factors put a child at risk of not reaping benefits of later education, leading to lower rates of employment in adult-life, and making it more challenging to achieve economic stability.

The important role parents and caregivers play in caring for and providing a rich and nurturing environment to children during these critical early years cannot be overstated. A myriad of studies have shown that parent and caregiver investment in a child's development leads to better physical, emotional, and cognitive growth. What was once based on inference has now been scientifically investigated through neuroscience research. Advances in brain imaging techniques have filled the gap in understanding how parent or caregiver input impacts the brain, and ultimately a child's skill formation. The evidence generated suggests the single most important component to healthy brain development is the relationship between a child and their caregiver, with warm and responsive interactions at the heart of that relationship.

Research from economics reveals that investment in the care and development of children during the early years makes economic sense, too. It not only leads to better outcomes for individual children and their families but impacts the society as a whole. In a series of groundbreaking studies, economist James Heckman has shown that investing in early childhood development, particularly for at-risk children, can significantly reduce social and economic costs to society in the long run and help break the cycle of intergenerational poverty. Investments in early childhood lead to long-term positive impacts on a child's education, health, social behaviors, and employment outcomes, thus reducing societal spending on remedial education, social services, health care expenses, and criminal justice. Further, his analyses suggest that high-quality early childhood programs, as we mentioned previously, can have a 13% return on investment for every dollar spent. A failure to invest in the care and well-being of our children has significant social and economic costs.

In short, early childhood development is a public health issue—one that merits a robust population-level health response targeting parent and caregiver investment in child development, beginning at birth.

The Vast Landscape of Early Childhood Programs in Developed and Developing Countries

The growing body of literature on the importance of early childhood has led to increased support for early childhood programs and policies worldwide. The United States established the Administration for Children & Families (ACF) in 1991 to promote the resilience, health, safety, and economic security of children, youth, families, and communities through the implementation of educational and supportive programs in collaboration with organizations at the state and local levels. In 2015, the United Nations General Assembly included early childhood as a target within the goal on education in its *Sustainable Development Goals (SDGs)*, a set of 17 interconnected goals designed to achieve a more sustainable future for all. Furthermore, one of the world's oldest and most-respected medical journals, *The Lancet*, published special issues devoted to early childhood in 2007, 2011, and 2017. As momentum among policy makers, researchers, and practitioners has accelerated, both developed and developing countries have made significant strides in strengthening their early childhood care and education programming.

Of course, that programming differs from country to country. Many developed countries, having already eradicated issues of chronic malnutrition, are now heavily focused on supporting children's cognitive and socio-emotional development. Developing countries are enhancing their programs to address all three simultaneously in response to research that shows strengthening the early childhood landscape requires an integrated, multi-dimensional approach focused on developing children's physical, cognitive, and socio-emotional development.

Reviewing the progress that all countries have made to improve child outcomes across all three dimensions tells one story: It is complex and challenging, even when resources are abundant. If we, as a society, want to improve outcomes for all children, it is imperative that we better understand the threats to scalability so that resources can be used strategically on interventions that will actually lead to desired outcomes. This is particularly true in countries with scarce resources and little to no infrastructure. In

this book, we draw upon evidence and lessons from early childhood programs in the United States and across the world to help develop a shared understanding of the challenges to scaling and advance the thinking of researchers, practitioners, and policy makers trying to improve outcomes for millions of children across the world.

How Early Childhood Programs Reach Children and Families

Irrespective of their location, early childhood programs around the world generally reach children through one of two delivery channels: (a) through services provided directly to children in the form of child care programs (center-based or home-based) and preschool programs or (b) through parent-focused interventions delivered in homes or in community or health-care settings.

Let us take the example of the United States to review the incredible diversity of early childhood care and education (ECE) programs¹ that cater to children under age 5. Of the nation's roughly 20 million children in this age group, a majority spend time in multiple care settings during a week—for example, children may spend time at a child care center on some days and be watched by a relative or neighbor on other days. The programs that serve them differ greatly in terms of the ages of participating children, the setting, funding mechanisms, and hours of operation, as well as quality. The National Survey of Early Care and Education (NSECE) 2014 Report estimates there are approximately 129,000 center-based programs in the United States that serve roughly 7 million children between birth and age 5. In its classification, NSECE includes many types of ECE programs including child care centers, Head Start, preschools, and pre-kindergarten classrooms in public schools. Per their estimates, 30% of centers exclusively serve children ages 3 to 5, 25% serve children birth to age 5, and another 27% serve children from birth to school-age. If we examine this further, it is apparent that there are more centers that serve

1

Notes

ECE programs are defined as those serving children age birth to five years and not yet in kindergarten.

the 3-5-year-old population than birth to 3; furthermore, there are more centers that serve 2-year-olds than those that serve children younger than 1 year. In terms of the organizational structure, as many as 52% of these centers are not-for-profit, whereas 32% are for-profit and only 16% are run by government entities. Sixty-eight percent of centers generate revenues through tuition and fees from parents and a majority of the programs (roughly 72%) receive some revenue from government sources.

Center-based child care programs are just one of many entities that deliver services directly to children in early childhood. There are also many entities that focus on parents. For example, there are a number of home visiting programs that support and empower pregnant women and families with young children—particularly those living in at-risk communities—with the knowledge, resources, and skills they need to thrive. Twenty-one home visiting programs in the United States meet the Department of Health and Human Services' criteria for an “evidence-based early childhood home visiting service delivery model” (Sama-Miller et al., 2019). These programs, which we know have positive impacts for vulnerable children and families, unfortunately reach only a tiny percentage of families. Estimates suggest that only 1.6% of families that stand to benefit from evidence-based home visiting were served by such programs in 2018 (Meisch & Isaacs, 2019). In other words, more than 17.5 million families with a child under age 6 are not receiving this essential service.

These numbers tell a story that is far too common: An unacceptably high number of children and families across the world do not receive the support they need to foster healthy brain development during the early years.

Why is Early Childhood So Susceptible to the Challenges of Scaling?

Despite the science that highlights the importance of nurturing relationships for children's development, why have we failed to provide most families with access to high-quality early childhood programs that offer or bolster such relationships? Why have these programs, which have demonstrated

evidence of effectiveness, not been scaled to the point where they serve the intended population? Of course, funding (or the lack thereof) is one of the biggest challenges. While the discussion on funding mechanisms and how best to allocate funds is outside the scope of this book, we hope that a deeper understanding of threats to scalability will go a long way toward not only developing programs that have evidence of effectiveness at scale, but will also help catalyze funding to further support such programs.

Funding aside, let's consider some of the challenges that pose a challenge to scaling in the early childhood field in the United States.

Early Childhood Programs are Implemented by a Complex, Fragmented System

During the first years of a child's life, children and families interact with multiple, fragmented systems that span the fields of health, family support and engagement, and early learning and development. There is no one universal system that exists for early childhood development, like the K-12 public school system, which children don't enter until age five or six. Consequently, families have to work through multiple systems to secure appointments and visit multiple offices to meet their children's needs. Besides being inefficient and an added burden on families, it puts the most vulnerable families and children at risk of falling through the cracks.

In addition to the challenges posed by the number of programs available, differences in governance structure, infrastructure, and resources available across states makes it incredibly difficult to effectively replicate a program across geographies and maintain results. Researchers and practitioners can design a highly effective intervention, but unless there is an equally robust infrastructure and governance structure to uptake the intervention, population-level impacts will remain an aspirational goal.

Lack of Universal Data System and Scalable Measures of Quality Leaves Many Flying Blind

Access to reliable and accurate data is critical to any scaling effort. But individual early childhood programs differ significantly in the kind of data and quality measures they collect or are required to

collect. Almost all programs have some administrative data available (demographic information on participating children and families, program attendance, staff qualification, caseload etc.), however, the depth and breadth of this administrative data varies significantly across programs and states. While there has been movement in states toward integrating data across early childhood programs to reflect all the services families may access and to measure long term outcomes, many of these efforts are still in the early stages.

While administrative data can provide valuable insight into a program, it doesn't paint a complete picture. The quality of an early childhood program depends on both structural quality (child-staff ratio, group size, training etc.) as well as process quality (i.e., the day-to-day interactions a child has with parents and caregivers). In fact, process quality is at the heart of a high-quality early childhood program and drives outcomes that are distal and appear years later. Measuring process quality isn't easy and is particularly challenging at scale. Measures used to study process quality, especially those used in research studies or ones that measure impacts on children's cognitive outcomes, are resource-intensive, time-consuming, and cannot realistically be used at scale.

In the absence of an integrated data system and scalable measures of quality, it is difficult to understand how programs perform relative to one another. This makes it especially difficult for decision makers to choose the right program for their communities, thus making effective scaling a challenge.

Workforce is Underpaid, Under-trained, and Susceptible to High Turnover

According to the National Survey of Early Care and Education (NSECE) 2013 Report, many early childhood teachers and caregivers have relatively low levels of educational attainment. This varies significantly for home-based teachers and providers versus center-based providers, as well as by the age group of children they serve. Among center-based teachers, only 45% of teachers of 3- to 5-year-olds had a bachelor's degree or higher and only 19% of those catering to 0 to 3-year-olds had the same. A recent Institute of Medicine & National Research Council report (2015) had strong recommendations to

require bachelor's degrees for lead teachers in order to gain parity with elementary school teachers, eliminate barriers for teachers to attain additional training and degrees, and pair knowledge with on-the-job coaching for ongoing professional development. At this stage, however, few states and few providers have been able to enact these recommendations. (Whitebook et al., 2018).

The NSECE survey also found that center-based early childhood teachers and caregivers were paid an average hourly wage of \$10.60, with those caring for infants and toddlers receiving lower wages than those serving 3- to 5-year-olds. It is hard to incentivize child care providers to invest in a bachelor's degree and then return to these low-paying roles. In fact, these dismal wages are among the most common reason for the high turnover among early childhood providers. The Center for the Study of Child Care Employment's 2014 report suggests a turnover rate of 15% among preschool and child care providers, the highest of any teaching profession (Whitebook et al., 2014).

Lastly, these wages are significantly lower than the median wage across all occupations (with only fast-food workers earning even lower wages). So it is no surprise that many teachers and caregivers are unable to meet their own basic needs and therefore participate in public assistance programs (Whitebook et al., 2018). There can be no bigger irony than this: We expect these caregivers to provide the highest-quality care to our children and to help those at-risk break free from the cycle of poverty, yet we refuse to compensate them adequately for that critically important work.

In order for evidence-based programs to produce their anticipated outcomes, programs must be implemented with quality and consistency. The early childhood workforce must be afforded the professional development and wages necessary to support this critical component of implementation.

How Do We Move Forward?

These are critical, thorny issues in the policy landscape that no individual stakeholder—researcher, practitioner or policy maker—can fix on their own. Instead, we must forge stronger research, practice,

and policy partnerships to tackle these complex issues. We face a collective and urgent need to understand the threats to scaling, because doing so will enable us to develop programs that work—and work within the constraints of relevant systems and contexts.

The question of scaling is a timely and critical endeavor for the TMW Center for Early Learning + Public Health, co-directed by two of us (Dana Suskind and John List). At the TMW Center, we develop and test interventions that place parents and caregivers at the center of their children's education. Our team has developed a suite of interventions designed to be embedded in existing health, education, and social systems, in order to meet parents where they already are. Ultimately, the TMW Center aspires to create a population-level shift in the knowledge and behavior of parents and caregivers, and optimize foundational brain development in children, particularly those born into poverty. To that end, we often wrestle with the question of scaling as we consider what it will take to adapt TMW interventions, which have been tested in small-scale studies, for implementation at the community level. How can we be sure of the impact at scale? Would the results of our interventions hold up in a different location or with a different population? What can we do to ensure they do?

The Science of Using Science: The Next Frontier in Evidence-Based Policy Making

Evidence-based programs were once an aspirational goal, then became the gold standard, and now they are the expectation. The research community has made significant strides in generating data and exploring intervention effects. And many researchers have also studied the challenges associated with scaling, particularly through the lens of implementation science. Yet, what has been lacking is a scientific understanding of how to make optimal use of the scientific insights generated. In particular, *how should we use experimental insights for policy purposes?*

In a new series of papers, two of us (John List and Dana Suskind), together with our colleague Omar Al-Ubaydli and Min Sok Lee, use a different lens to examine this issue: economics. We have developed an

economic model to explore the phenomenon by which the magnitude of a treatment effect changes when an intervention moves from research setting to population-wide implementation. We've dubbed this the *scale-up effect*.

The scale-up effect revolves around several important questions: When we scale an intervention to broader and larger populations, should we expect the same level of efficacy that we observed in the small-scale setting? If not, then what are the important threats to scalability? What can the researcher do from the beginning of their scholarly pursuit to ensure eventual scalability?

Providing answers to such questions is necessary because understanding when and how our experimental insights scale to the broader population is critical to ensuring a robust relationship between scientific research and policy making. Without such an understanding, empirical research can quickly be undermined in the eyes of the policy maker, broader public, and the scientific community itself.

Our economic model is an attempt to provide answers to some of these questions and to raise additional questions that advance the *science of using science*. We hope that through this new economic lens, we can provide a framework to understand the scale-up effect—the underpinning threats to scalability, including incentives and challenges stakeholders face in developing and testing evidence-based programs—and that this deeper understanding will allow us to develop solutions to attenuate the scale-up effect.

Components of the Scale-Up Effect (AKA Threats to Scaling). The scale-up effect, as defined in the economic model, refers to the net changes in benefits and costs of a program when it is implemented at scale. This definition differs from the more commonly discussed voltage drop phenomenon, which considers only changes in benefits—but not costs—brought about by the scaling of a program. The

economic model identifies four categories of threats that are key to understanding the scale-up effect (the economic model and the individual categories are explored in detail in Section 2 of this volume):

- **Inference** - Threats to scalability emerge when an intervention tested in a small-scale randomized controlled trial suffers from an inference problem, such as a false positive, and then is implemented at a larger scale. In other words, the intervention is scaled even before there is sufficient evidence of efficacy.
- **Properties of the population** - Threats to scalability emerge when an experiment's participants are not representative of the policy-relevant population. A non-representative participant pool can be caused by several factors, including the fact that in a competitive scientific marketplace, researchers may be more likely to choose a subject pool with a greater chance of showing a treatment effect than would a random sample. Similarly, when selecting a program, policy makers and practitioners may overlook the fact that the program was only tested with one population or assume it will apply to their target population as well.
- **Properties of the situation** - Threats also emerge when the experimental situation is not representative of the policy-relevant situation. Properties of the situation include specifics of the program, correct dosage, correct delivery, and implementation costs.
- **Spillover** - Spillover refers to the phenomenon where individuals who participate in a program may interact with individuals who do not participate in the program, potentially sharing information that leads to unanticipated consequences (positive or negative) and makes it harder to isolate and evaluate the true impact of the program.

How Does this Book Advance the Field's Understanding of Scaling?

This book aims to provide a robust picture of the science of scaling through the lens of early childhood. The book begins by orienting readers to why this topic is so critical, then transitions to a deep discussion of the economic model of scaling and all of its individual components, as well as the

infrastructure necessary to support scaling. The last section of the book addresses important factors that can affect scaling but are not explicitly part of the model, such as political will and research-practice partnerships. The chapters' authors bring perspectives from multiple disciplines (e.g., economics, psychology, and public health) and apply multiple lenses (e.g., research, policy, and practice) to the topic. Each of the three sections, and the chapters within, are summarized below.

The Science of Early Childhood and Complexities of Scaling

The first section of the book lays the foundation for why we are focusing on early childhood as a way to explore challenges to scaling. It provides an overview of the key principles of the early childhood field. Though each chapter focuses on a particular principle, taken together, they demonstrate why it is so critically important that we learn to scale evidence-based early childhood interventions (practices, programs, and policies)—and why it's particularly challenging to do so.

Chapter 2 by Kim Noble highlights the critical importance of early childhood experiences on a child's developmental trajectory and brain development. It highlights the body of research showing that children's earliest experiences—centered on the home environment and caregiving practices—play a key role in shaping brain development. It delves into the relationships between socioeconomic disadvantage and early adversity and brain development, and explores how early life experiences can serve as targets for interventional and preventive strategies to support healthy development. Supportive environments and caregiving have been shown to create lasting protective factors against early adversity and the effects of poverty for children.

Chapter 3 by Susan Mayer, Rohen Shah, and Ariel Kalil then identifies why selecting and implementing evidence-based early childhood programs may not be as logical and linear as one would expect. The influence of cognitive biases, such as confirmation bias, status quo bias, or bandwagon bias, may preclude researchers, decision makers and practitioners from selecting the right programs and

implementing those programs with quality and fidelity. Confirmation bias can affect how evidence is received and interpreted, or whether a program or particular study is funded. Status quo bias highlights the challenge of making a proactive change to implement an evidence-based program. Finally, bandwagon bias highlights the strong influence of others on our decisions. This bias may lead to ineffective programs scaling because others have decided to do so rather than based on strong research. This chapter highlights not only a critical piece of the decision-making process that might lead to scaling but also a gap in the existing research on ways to mitigate the influence of cognitive biases on scaling.

Chapter 4 by Lisa Gennetian adds to the discussion on bias and decision making by introducing the connection between behavioral economics and parent engagement in early childhood programs. This framework is critical to the discussion of scaling. First, without strong parent engagement in programs, the desired outcomes for children are unlikely to be realized. Currently, parent engagement in programs is a challenge in the field. Second, the framework recognizes that the context of living in poverty places additional cognitive demands on parents that make it difficult for them to have high engagement, meaning programs need to be designed with behavioral economic principles in mind. This can mean, for example, offering parents structured choices or leveraging social norms by having parents recruit other parents. The chapter concludes with a few examples of the use of a behavioral economic framework to support parent engagement in early childhood programs.

Finally, Chapter 5 by James Heckman, Bei Liu and Jin Zhou demonstrates the incredible economic benefits that would be generated by the successful scale-up of evidence-based programs in early childhood. In order to reap those benefits, the chapter argues, programs must be designed with a clear understanding of the fundamentals of children's skill formation, as well as the program mechanisms driving positive outcomes and the context in which potential scaling will occur. The chapter examines these topics using the examples of four home-visiting programs from around the world, including one (China REACH) that was a scaled replication of a successful experimental program.

Understanding the Threats to Scalability

The second section of the book is dedicated to identifying and understanding the threats to successful scaling of evidence-based programs. Why is it that so many programs that are found to be successful in small-scale research studies fail to deliver the same level of impact when expanded to a broader population of participants? The economic model introduced above is a framework for studying the science of scaling, and each of the model's pillars will be examined in dedicated chapters within this section of the volume. The section starts with a case study about a highly impactful and oft-cited early childhood program in Jamaica and traces its scale-up journey to larger populations in Colombia and Peru.

In Chapter 6, together with Omar Al-Ubaydli and Min Sok Lee, two of us (Dana Suskind and John List) introduce the science of using science (also referred to as the economic model throughout the book), which combines economics and the experimental method to inform the conversation about scaling evidence-based programs. Though the overall book takes a broad view, the economic model focuses on the role of the researcher in the knowledge-creation system. The components of the model were already described above and are further discussed in this introduction. One notable aspect of this chapter is the use of a fictional case study to ground the concepts from the model and illustrate how they might play out in a community. This example is sure to stick with readers as they progress through the next four chapters, which dig deeper into each component of the model.

Chapter 7 by John Ioannidis, Zacharias Maniadis, and Fabio Tufano discusses the scientific principles and methods that can help decision makers know if evidence from a promising program is ready to be acted upon. The chapter starts with examples of three early childhood programs that were found to be effective at a relatively small scale but failed to demonstrate the same impacts when replicated or scaled up. In some cases, the original research mistakenly concluded the program was effective when it was not, which led to the inability to replicate evidence. In others, it was because the original research was

simply not designed with scaling in mind. All of these challenges raise the importance of two factors to consider prior to scaling. First, the need for researchers to be clear with potential users of their research about how much the findings can be applied to other populations or contexts. Second, researchers need to be clear about how confident they can be in the impacts obtained in a given trial. Often, policy makers and practitioners are enthusiastic about a program and jump to implementation before researchers are confident it has been tested in a rigorous way or that the findings can be replicated. The authors introduce five principles to help assess the scalability of evidence—priors, statistical power, bias and conflict of interest, competition and publication bias, and replication approaches—and conclude with concrete recommendations for decision makers who must choose whether and when to scale-up a particular program.

The theme of external validity is an important parallel to the internal validity described in the Ioannidis and Al-Ubaydli chapters. One of the pillars in the economic model is properties of the population, or the degree to which a study's participants accurately represent the population to which the program is being applied. When study participants are not representative, we have less confidence that the impacts will transfer to a broader population when the program is scaled. In Chapter 8, Jonathan Davis, Jonathan Guryan, Kelly Halberg, and Jens Ludwig discuss the importance of generalizability and scalability to ensure programs are designed for implementation at scale. The chapter provides a framework for assessing the ability of research to achieve generalizability. The chapter provides recommendations for researchers to enhance the external validity of their research, including within four specific research designs.

Properties of the situation is another important element of the economic model. Chapter 9 by EB Caron, Kristin Bernard, and Allison Metz discusses one particular aspect of properties of the situation: the role of administration quality as a key factor in ensuring program effectiveness at scale. Achieving administration quality requires strong infrastructure, a qualified workforce, and ongoing training and

supervision. The authors describe these concepts and provide a case study of a home visiting model, Attachment and Biobehavioral Catchup (ABC), and the efforts to move ABC from a university-developed evidence-based model to scale in New York City. The chapter discusses the challenges faced in that process and provides recommendations for researchers, policy makers and practitioners based on the lessons learned.

Finally, the economic model highlights the innovative idea of spillover effects. In Chapter 10, Fatemeh Momeni and Daniel Tannenbaum discuss the central importance of understanding and measuring these concepts to support scaling. Spillover occurs when participants in a program or non-participants experience different outcomes, either positive or negative, due to their interactions with each other. In order to study the scaling of programs, the field must do a better job of measuring spillovers to truly understand the potential impact of a program on individuals and systems at scale. The chapter raises the possibility that we are underestimating the effects of currently scaled programs because this factor is not well understood or measured.

In the next chapter, Chapter 11, M. Caridad Araujo, Marta Rubio-Codina, and Norbert Schady describe how one early childhood program—a home-visiting intervention originating in Jamaica—has been expanded and implemented at different scales. After an initial study found the Jamaica program to have impressive positive impacts, it was adopted in multiple other locations. This chapter focuses on two locales in particular, Colombia and Peru, which both adapted the program to their local context. By exploring the differences in context, implementation, and ultimately, results, this chapter provides an excellent case study to understand the concepts and threats to scalability discussed in chapters 7-10. The themes raised in this chapter reappear throughout the book.

Sophia Pappas shares her perspective and insight in a short commentary that ties the content of these chapters directly to real-world early education contexts. She draws on her rich experience in

scaling pre-K in New York City and now leading the BEST initiative, a place-based early childhood strategy in Tulsa, Oklahoma. Her commentary discusses central factors such as the importance of timing and trust in research-practice partnerships and the need for dynamic and iterative relationship building between partners. Pappas raises nuanced factors at play in the threats to scalability that are embedded within research-practice partnerships.

The section rounds out with Chapter 12 by A. Mushfiq Mobarak and C. Austin Davis about building a research agenda for scale. Using the example of the Yale Research Initiative on Innovation and Scale (Y-RISE) as a case study, the chapter explores ways to shift researchers' attention and allocation of time toward exploring key questions in the science of scaling. It focuses on three activities in particular: a) providing start-up resources, b) organizing research networks, and c) coordinating research through research initiatives. The chapter discusses how these activities can support and facilitate research on scalability. The chapter also raises two subjects that are touched on throughout the book: the importance of research-practice partnerships and long-term, ongoing evidence building as critical to effective scaling.

Charting a Path Forward

The science of scaling is still very new. Consequently, this book raises more questions than answers, and at the same time suggests some broad approaches to move the field forward. The last section of the book is dedicated to charting the path forward; it provides an on-ramp for addressing the many complex issues raised in the earlier sections of the book and a vision for the long-term journey to scalable evidence. The chapters focus on critical factors in the context of scaling, including program design, system supports, partnerships, continuous quality improvement and measurement, cost analysis, and policy and political will.

In the first chapter in this section, Chapter 13, Aaron Lyon asserts that many evidence-based programs are not designed with scaling in mind, discusses the consequences of that shortcoming, and identifies potential solutions. He explains how programs designed under ideal circumstances often run into implementation challenges when scaled in the field. The economic model in Chapter 6 (Al-Ubaydli et al.) suggests the use of technology as one potential solution to this issue. In addition to technology supports, Lyon recommends leveraging the techniques of human centered design to ensure that programs are designed with the end-user in mind, prior to testing. The chapter demonstrates how this forward-looking approach to design will reduce the overall costs of scaling.

Chapter 14 by Elizabeth Stuart focuses on how well impact evidence from research applies to other situations and populations. The chapter begins by demonstrating how the current evidence from impact evaluation may not, as assumed, generalize to new locations. The chapter describes threats to generalizability to new populations or settings and presents design and analysis strategies to address this issue. This chapter highlights the importance of challenging the assumptions built into our current system of evidence building—taking one or two randomized trials in specific locations and expecting those effects to be found at scale in more diverse settings and with different populations. This chapter is an excellent companion to the next chapter by Chambers and Norton.

In Chapter 15, David A. Chambers and Wynne E. Norton argue that sustaining impact after scaling requires continuous collection of data. This will help ensure expected impacts are achieved and inform any necessary refinement of the intervention. At scale, interventions should be allowed to evolve to address new learning and address changes in the implementation environment. The chapter discusses considerations for how to sustain impact based on these new assumptions and what tools can support this process.

Another consideration when it comes to scaling is how to measure success of the scaled-up program. The measures that are often used in clinical trials are not the same as those used to study programs at scale. Measures in small-scale trials are often more time- and labor-intensive and more sensitive to change, which makes them difficult to use at a population level. Data on the outcomes that programs aim to target are often not even collected in administrative data systems. These disconnects make it hard to measure the magnitude of effects and the cost-benefit ratio of scaled programs. Chapter 16 by Scott McConnell and Howard Goldstein describes these and other tensions related to measurement at scale and, utilizing design principles of measurement, provides recommendations for the future.

The economic model at the heart of this book emphasizes the use of benefit-cost analysis as a critical component of the science of scaling. However, the available research on most evidence-based early childhood programs and policies lacks cost information and thus, benefit-cost analysis. When cost analyses are available, they are often on tightly controlled clinical trials, which may limit generalizability to the actual benefits and costs at scale. Meryl Yoches Barofksy, Rachel Herzfeldt-Kamprath, and Kyle Peplinski's commentary highlights the challenges of measuring costs on the ground in both Head Start and home visiting programs. Their commentary presents some of the challenges to collecting cost data that may present barriers to better understanding costs at scale and measuring the return on investment of specific programs at scale.

Another key ingredient for successful implementation at scale is the systems that surround programs and the extent to which they support scaling. Chapter 17 by Melissa Lim Brodowski and Sandy Naoom focuses on the contexts within early childhood programs and the systems in which they operate. Their chapter examines elements of system-level infrastructure that support implementation and scale-up of evidence-based programs. Using examples from the Early Head Start-Child Care Partnership program, which aimed to address parent needs by partnering a federal early childhood program (Early

Head Start) with existing, community-based independent child care centers and creating a new context, they highlight key elements of context and systems that facilitated successful scaling of this national program in a selection of states.

One of the critical infrastructure challenges faced in scaling evidence-based programs is related to the availability of a qualified workforce. Implementing an evidence-based program accurately (i.e., with fidelity) is challenging. It requires training and support. Whether or not a program is implemented with fidelity, and whether or not it realizes economies of scale, largely rests on the shoulders of the workforce. In Chapter 18, Debra Pacchiano, Maia Connors, Rebecca Klein, and Kelly Woodlock discuss some of the complex challenges faced by the early childhood field in identifying and retaining a qualified workforce. Then they present a recommended approach to supporting the development of quality leadership, which can, in turn, support scaling. Their framework considers the need to establish infrastructure and consider sustainability from the start of any scaling effort.

Recent examinations of effective strategies for translating research into policy and practice highlight the critical role of research-practice partnerships. Chapter 19 by Samantha Carter, Iqbal Dhaliwal, Samantha Friedlander, and Claire Walsh explores the importance of such partnership using as a case study the Abdul Lateef Jameel Poverty Action Lab (J-PAL)'s experience bringing evidence-based programs to scale through complex partnerships between policy makers, practitioners, researchers, and funders. Their chapter describes these collaborations and the lessons learned through them, including the importance of policy windows, complementary data and evidence, a culture of evidence, and intermediary organizations. The chapter highlights examples from scale-up work internationally.

While many of this book's chapters indirectly focus on evidence-based *programs*, the book is inclusive of practices, programs, and policies. Chapter 20 by Cynthia Osborne focuses on the latter. Policies differ from programs in that they are typically adopted through legislative or regulatory channels and apply broadly to a population based on stated eligibility criteria. This chapter explores the unique

challenges in identifying which state-level policies are effective, and provides recommendations to researchers and state policy leaders on how we can develop an evidence-driven policy agenda over time.

Without political support, it would be impossible to scale up evidence-based programs given the central role policy makers play in funding research and programs, setting requirements, enforcing accountability, and ensuring sustainability. The book ends with a case study to highlight political will as a factor not in the economic model but central to the topic of scaling. While not a focus of this book, political will is critical to decisions on scaling evidence-based programs and the population-level impacts desired. Chapter 21 by Mary Young and Osmar Terra describes the components necessary to build the political will to scale programs. The chapter uses the scale up of Brazil's national home visiting program as a case study to bring to life that process, including the pressure to scale quickly and the importance of political champions and the sustainability of political and financial support.

Across these 21 chapters, several overarching themes emerge that are central to understanding and addressing the scale-up effect. Understanding these themes is critical to further refining the economic model and creating practical solutions to the challenges of scaling over time.

Overarching Themes Across Chapters

Across all of the chapters in this book, some clear themes emerge. These themes represent some of the important next steps in refining the science of scaling and addressing the scale-up effect. Some of the themes have to do with quality of the available evidence, such as the importance of ensuring external validity, measuring spillover effects, and generating better cost data. The themes around inference reveal the need for strong partnerships with the research community to make change. The economic model includes suggestions around the incentive structure to ensure the changes necessary in research can and do occur.

One overwhelmingly consistent theme is the need for strong, ongoing partnerships between research, practice, and policy. None of the discussion in this book is relevant or actionable in the absence of such partnerships. Fortunately, past research has identified three components of effective partnerships: mutualism, commitment, and trust (Tseng et al., 2017), which complement the economic model and can help guide the creation of partnerships to support scaling.

The economic model discusses the critical importance of designing studies that are more relevant to the real world. The first component of effective partnerships, mutualism, supports this goal. Mutualism means deep authentic collaboration where researchers and practitioners jointly define the research agenda. Commitment is relevant because of the long-term investment required of partners. Research-practice partnerships in pursuit of scaling require iterative work to understand key problems and needs. Finally, trust is a key requirement that enables each partner to proactively anticipate each other's needs and interests. The work of J-PAL and others demonstrates these key principles and how they can support the development of research to support effective scaling. While not an explicit part of the economic model, in order to address some of the components that are included, the formation and continuation of RPP is a necessary, but not sufficient requirement.

Another central theme is the importance of infrastructure, including data systems, measurement systems, workforce, leadership, and political will. Infrastructure is a key subject studied throughout implementation science. It is critical that all infrastructure components are embedded within research to support effective scaling. Thus, it is clear that partnerships between researchers in the implementation science field and those conducting the research and practice described in the economic model will be critical for success.

Currently, federal and state policy making has a focus on evidence-based policy. This presents an exciting opportunity to build momentum around the need to address the issues raised in this book—and

to take the steps necessary to do so. The evidence-based policy movement has shown signs that it's ready to tackle the challenges discussed here. For example, evidence requirements are being included in the design and funding of new programs. The Family First Prevention Services Act of 2018 requires replicated impact evidence tested under real-world conditions to qualify for the highest evidence standards, tied to the highest reimbursement rates for states. In another example, the Foundations of Evidence Based Policy Making Act of 2018 includes a focus on building and using data to inform policy and includes means of strengthening evidence-building capacity, such as the requirement of agency learning agendas. These are long-term research plans similar to those described in the Y-RISE case study, which help identify gaps in evidence and address them in a multi-tiered evidence-building strategy. The Act also aims to improve federal data capacity to make data more accessible and easily used in evidence building. This Act complements work at the state level around integrating data across government programs such as education, health, and early childhood (King et al., 2018). Integrated data systems allow for continuous quality improvement and measurement of long-term outcomes of evidence-based programs.

Those of us concerned with tackling the scale-up effect cannot operate in a silo. We can and should leverage existing movements to continue to improve the available evidence—and to more strategically address the scale-up effect in order to achieve population level-change. This is a marathon, not a sprint. We will not make such gains immediately, but by working together, we will make critical progress.

Recommendations

This book concludes with recommended actions to continue to make progress toward effective scaling for population-level improvement of outcomes. Our recommendations are aimed at several critical stakeholder groups, each of which can do its own part: researchers; agency heads; policy makers; and funders.

The charge for researchers is a clear: Design research that utilizes more representative populations and situations and measure critical characteristics of experiments, such as cost and spill-over effects. Researchers need to implement open science principles such as pre-registration and analysis plans to address issues of null effects and allow for replication. Researchers also need to build and flex their partnership muscles in order to co-create evidence with policy makers, and thus make it more actionable.

The recommendations for the leaders of organizations that select and implement evidence-based programs, such as state agency directors, is to demand and use contextual information when deciding whether to implement a specific intervention. They must then assess the relevance of the research for their own organization and act based on the available evidence. Organizational leadership also needs to build measurement and continuous quality improvement systems that support ongoing learning and quality implementation. All of these recommendations are grounded in building internal capacity to understand effective implementation, the use of data for decision making and oversight, and skills to work with research partners or intermediaries.

The recommendations for policy makers build on those for organizations. They include understanding if and how available research is applicable to the policy need at hand, understanding the multiple factors of implementation necessary for success (e.g., workforce, data, replication research) and providing the funding to support their creation and maintenance. Finally, policy makers, too, have a role to play in the creation and maintenance of research, practice, and policy partnerships to shape and interpret research, identify key questions and outcomes, and support the conditions necessary for effective scaling.

Finally, the recommendations for funders include designing funding priorities and requirements to emphasize evidence-building for scaling, including replication studies, representative populations and

contexts, measuring cost and implementation, and open science practices. For those funders investing in program services, the recommendations call for using evidence that meets the scaling criteria and, similar to policy makers, supporting the funding of the necessary infrastructure as much as the specific program.

These recommendations provide a roadmap to guide the important work of the science of scaling. While the identified stakeholders may not enact all recommendations immediately, making progress toward them has the potential to address the critical need of mitigating the scale-up effect.

References

- Al-Ubaydli, O., Lee, M. S., List, J. A., & Suskind, D. (in press). The Science of Using Science: A New Framework for Understanding the Threats to Scaling Evidence-Based Policies. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.
- Araujo, M. C., Rubio-Codina, M., & Schady, N. (in press). 70 to 700 to 70,000: Lessons from the Jamaica Experiment. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.
- Barofsky, M. Y., Herzfeldt-Kamprath, R., & Peplinski, K. (in press). Commentary: The Challenge of Measuring Costs at Scale. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.
- Brodowski, M. L., & Naom, S. (in press). Enabling Contexts to Support Scale-Up: Lessons from Early Head Start- Child Care Partnerships. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.
- Caron, EB, Bernard, K., Metz, A. (in press). Fidelity and Properties of the Situation: Challenges and Recommendations. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

Carter, S., Dhaliwal, I., Friedlander, S., & Walsh, C. (in press). Forging Collaborations for Scale: Catalyzing Partnerships Among Policymakers, Practitioners, Researchers, Funders, and Evidence to Policy Organizations. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

Chambers, D. A., & Norton, W. E. (in press). Sustaining Impact after Scaling Using Data and Continuous Feedback. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

Davis, J., Guryan, J., Hallberg, K., Ludwig, J. (in press). Studying Properties of the Population: Designing Studies that Mirror Real World Scenarios. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

Gennetian, L. (in press). How a Behavioral Economic Framework Can Support Scaling of Early Childhood Interventions. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

Heckman, J. J., Liu, B. & Zhou, J. (in press). The Economics of Child Development with an Application to Home Visiting at Scale. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

Institute of Medicine and National Research Council (2015). *Transforming the Workforce for Children Birth Through Age 8: A Unifying Foundation*. The National Academies Press.
<https://doi.org/10.17226/19401>.

Ioannidis, J. PA, Maniadis, Z., Tufano, F. (in press). When is Evidence Actionable? Assessing Whether a Program is Ready to Scale. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

King, C., Perkins, V. Nugent, C., and Jordan, E. (2018, September). *2018 State of Early Childhood Data Systems*. Child Trends.
<https://www.childtrends.org/wp-content/uploads/2018/09/ECDC-50-state-survey-9.25.pdf>

Lyon, A. R. (in press). Designing Programs with an Eye Toward Scaling. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

Mayer, S., Shah, R., & Kalil, A. (in press). How Cognitive Biases Can Undermine Program Scale-Up Decisions. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

McConnell, S., & Goldstein, H. (in press). Measurement Built for Scale: Designing and Using Measures of Intervention and Outcome that Facilitate Scaling Up. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

Meisch, A., & Isaacs, J. (2019, December). Exploring home visiting's unmet need: Comparing who could benefit to who is served. *National Home Visiting Resource Center Data in Action Brief*. James Bell Associates and the Urban Institute.

Mobarak, A. M., & Davis, C. A. (in press). A Research Agenda Built for Scale. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

Momeni, F., & Tannenbaum, D. (in press). Unintended Consequences: Spillover and General Equilibrium Effects. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

National Survey of Early Care and Education Project Team. (2013). Number and Characteristics of Early Care and Education (ECE) Teachers and Caregivers: Initial Findings from the National Survey of Early Care and Education (NSECE). OPRE Report #2013-38, Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

National Survey of Early Care and Education Project Team (2014). Fact Sheet: Characteristics of Center-based Early Care and Education Programs. OPRE Report No. 2014-73b, Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
<http://www.acf.hhs.gov/programs/opre/research/project/national-survey-of-early-care-and-education-nsece-2010-2014>.

Noble, K. G. (in press). Early Childhood: The Opportunity to Untap Human Potential. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

Osborne, C. (in press). Process to Identify Effective Policies to Strengthen the Prenatal-to-Three System of Care. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

Pacchiano, D., Connors, M., Klein, R., & Woodlock, K. (in press). Embedding Workforce Development into Scaled Innovations to Prevent Declines in Administration Quality. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

Pappas, S. (in press). Commentary: Real-world Application and Understanding of these Threats. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

Sama-Miller, E., Akers, L., Mraz-Esposito, A., Coughlin, R., & Zukiewicz, M. (2019). *Home Visiting Evidence of Effectiveness Review: Executive Summary September 2019* (Rep. No. #2019-93).

<https://www.acf.hhs.gov/opre/resource/home-visiting-evidence-of-effectiveness-review-executive-summary-brief-september-2019>

Stuart, E. A. (in press). Accounting for Differences in Population: Predicting Intervention Impact at Scale. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.

Tseng, V., Easton, J. Q., & Supplee, L. H. (2017). Research-Practice Partnerships: Building Two-Way Streets of Engagement. *Social Policy Report*, 30(4), 1–17.
<https://doi.org/10.1002/j.2379-3988.2017.tb00089.x>

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. Center for the Study of Child Care Employment, University of California, Berkeley

Whitebook, M., McLean, C., Austin, L.J.E., & Edwards, B. (2018). Early Childhood Workforce Index – 2018. Center for the Study of Child Care Employment, University of California, Berkeley. <http://cscce.berkeley.edu/topic/early-childhood-workforce-index/2018/>.

Young, M., & Terra, O. (in press). Building Political Will. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.